Quantification of Inter- and Intra-Tumor Heterogeneity Using Medical Imaging and Its Implication on Response to Radiotherapy in Head and Neck Cancer

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Marta Bogowicz

aus

Polen

Promotionskommission

Prof. Dr. Martin Pruschy (Leitung der Dissertation, Vorsitz)
Prof. Dr. Matthias Guckenberger
PD Dr. Oliver Riesterer
Prof. Dr. Uwe Schneider
Dr. Stephanie Tanadini-Lang
Prof. Dr. Jan Unkelbach

Zürich, 2017

Abstract

Cancer is a heterogeneous disease, showing intra- and inter-tumor genetic and phenotypic variability [1]. This variability translates to differential radiosensitivity and in consequence differential response to radiotherapy. Head and neck squamous cell carcinoma (HNSCC) accounts for around 5-10% of new cancer cases in developed countries [2]. It shows a heterogeneous response to radiochemotherapy with loco-regional control and 5 years overall survival ranging from below 50% to 80%. Few molecular factors were linked to outcome prognosis in HNSCC, for example human papilloma virus (HPV) infection. However, tissue-based biomarkers from tumor biopsies may not account for intra-tumor heterogeneity [3, 4]. This PhD project aims to identify new tumor phenotypes in HNSCC related to worse prognosis of treatment outcome using medical imaging techniques, which provides a 3D surrogate of tumor biology. Tumor density, metabolism and perfusion were studied in respect to different HNSCC subtypes and radiochemotherapy outcome. A quantitative and comprehensive image analysis method, radiomics, was used to link intra-tumor heterogeneity and treatment outcome. Radiomics comprises four types of descriptors: shape, intensity, texture and filter-based. It not only quantifies general properties of a tumor, for example higher metabolic activity, but also provides information about the intra-tumor heterogeneity.

In the first subproject, I have analyzed tumor perfusion, metabolism and their correlation in subgroups of HNSCC based on: tumor subtype (oropharynx, hypopharynx, larynx and oral cavity), tumor stage (T1/T2 vs T3/T4) and HPV status. Computed tomography perfusion (CTP) and 18F-fludeoxyglucose positron emission tomography (18F-FDG PET), from 41 HNSCC patients were analyzed. Three perfusion parameters: blood volume (BV), blood flow (BF) and mean transit time (MTT), were computed. Difference in perfusion parameters between the gross tumor volume (GTV) and its surrounding tissue were investigated. Tumor subgroups related to worse prognosis (T3/T4 and HPV negative) showed increased BV and MTT in comparison to surrounding healthy tissue. Additionally, I have shown that the correlation of FDG uptake and perfusion is tumor subgroup dependent. I have observed positive correlation only for HPV positive ($r = 0.86$, $p = 0.04$) and oropharyngeal ($r=0.63$, $p = 0.05$) cancer.

CTP consists of repeated CT scans and is thus dose intensive. I have performed a separate study using Alderson phantom to adapt our clinical CTP head and neck protocol. The endpoint was a decrease in delivered dose and maintenance of image quality. Our standard protocol on GE revolution CT is 100 kV, 80 mAs, 5 mm slice thickness and filtered back projection algorithm. I have adapted the percentage of an adaptive statistical iterative reconstruction (ASiR), slice thickness, tube current and voltage. The signal to noise ratio was measured in 7 predefined regions of interest and the effective dose was estimated using thermoluminescent dosimeters. The optimized protocol used 80 kV, a tube current adapted based on anatomy from 15 to 80 mAs, 2.5 mm slice thickness and 50% of ASiR reconstruction. The effective dose was decreased by factor of 2 whereas the image quality was maintained.

In the second part of the thesis, I have investigated radiomics for its ability to predict treatment outcome and its correlation to tumor biology. An in-house radiomics software implementation was developed in Python programing language (v 2.7). Most of the radiomics studies are performed using in-house implementations or open source codes and the implemented workflows are currently not fully standardized. Therefore, I have validated my implementation against implementation from MAASTRO clinic, Maastricht, the Netherlands. I have also used both implementations to train local

tumor control models based on 18F-FDG PET imaging 3 months post-radiochemotherapy (128 patients). Only 80 out of 649 radiomic features, available in both implementations and based on the same mathematical definition, were reproducible between the implementations (intraclass correlation coefficient ICC > 0.8). In the univariate Cox regression feature's prognostic power depended strongly on used implementation. The main causes of irreproducibility were differences in contour mask creation, translation of bin size to filtered images, and type of the used transform decimated vs undecimated wavelet transform.

In another radiomics robustness study I have investigated the stability of radiomic features in respect to different CTP calculation factors. Some of the CTP calculation factors are difficult to standardize (arterial input function definition and noise threshold in the calculation) and thus should be considered before linking CTP radiomics with clinical outcome. I have analyzed CTP scans in lung (n = 11) and head and neck cancer (n = 11). 255 out of 945 CTP radiomic features were stable in both tumor sites in respect to artery contouring and noise threshold. Among them, I have identified 10 groups of radiomic features, after the correction for inter-features correlations and correlation to tumor volume. These features should be further tested for their prognostic power.

In the prognostic modeling, I have investigated the link between local tumor control and radiomics in HNSCC based on contrast-enhanced CT and 18F-FDG PET pre-treatment imaging. I have used two cohorts of patients: retrospective for models training (n > 90 patients) and prospective cohort from institutional phase II study with a standardized imaging protocol for models validation (n > 50 patients). I have observed that tumors more heterogeneous in CT density were at higher risk for tumor recurrence. This model had a higher prognostic power than model incorporating clinical prognostic factors (tumor stage, volume and HPV status) or combination of CT radiomics and clinical factors, concordance index (CI) in the validation cohort $CI_{radiomics} = 0.78$, $CI_{clinical} = 0.73$ and $CI_{combination} = 0.76$. In a follow-up study, I have investigated whether the inclusion of metabolic information can further improve radiomics-based local tumor control modeling. I have observed that round tumors (based on 18-FDG PET autosegmentation) with a focused region of high FDG uptake surrounded by a rim of low FDG uptake were linked with better prognosis. However, this model did not outperform the CT based model. In the validation cohort evaluated in this study, both models achieved CI around 0.7. Also the combination of PET and CT radiomics did not improve the predictions. Nevertheless, the PET radiomics model showed a better calibration, which may be linked to the presence of metal artifacts in CT in head and neck region.

To link the abstract radiomic features with tumor biology, I have correlated CT radiomics with HPV status. I have observed that tumors more homogenous in CT density tend to be HPV positive. Although, this signature (set of radiomic features) has a similar interpretation to local tumor control signature, it comprised different features and the signatures were not correlated with each other. For example local tumor control CT radiomics model was also prognostic in HPV negative subgroup of patients.

In summary, I have shown that biological information can be recovered even from simple morphological imaging (CT). Additionally, I have identified imaging signatures, based on differences in perfusion between tumor and its surrounding as well as CT and PET radiomics, which were linked with worse outcome prognosis. These signatures need to be further validated in an external cohort of patients and treatment intensification options for worse prognosis groups have to be defined.

Zusammenfassung

Krebs ist eine heterogene Erkrankung, die sich durch ihre genetische und phänotypische Variabilität auszeichnet [1]. Diese Heterogenität führt zu einer differentiellen Radiosensitivität und in der Folge zu einem sehr unterschiedlichen Behandlungserfolg der Strahlentherapie. Etwa 5-10% aller Krebserkrankungen in den Industrieländern werden durch das Plattenepithelkarzinom im Kopf-Hals-Bereiches (HNSCC) verursacht. Bei dieser Tumorart ist der Behandlungserfolg auf eine Radio-Chemotherapie sehr heterogen mit lokoregionären Kontrolle und 5-Jahres Überlebungsraten zwischen unter 50% bis 80%. Es konnten bisher nur wenige molekulare Faktoren festgestellt werden wie beispielsweise eine Infektion mit dem humanen Papillomavirus (HPV) die einen Einfluss auf die Prognose beim HNSCC haben. Jedoch können viele der Gewebe-basierte Biomarker aus Tumorbiopsien die intra-tumorale Heterogenität nicht richtig erfassen. Das Ziel dieser Doktorarbeit war es neue Tumor-Phänotypen für das HNSCC zu identifizieren, welche mit einer guten bzw schlechten Prognose assoziiert sind. Dazu wurden medizinischen Bildgebungsverfahren als 3D Surrogat für die Tumorbiologie verwendet. Tumordichte, Metabolismus sowie Perfusion wurden unter Berücksichtigung der unterschiedlichen HNSCC Subgruppen und Radio-Chemotherapie Ergebnissen untersucht. Die Heterogenität im Tumor wurde mit dem Behandlungsergebnis korreliert mit Hilfe einer quantitativen mathematischen Bildanalyse-Technik genannt Radiomics. In Radiomics Analysen werden vier Arten von Deskriptoren bestimmt: Form, Intensität, Textur sowie filter-basierte Features. Zum einen quantifiziert Radiomics die allgemeinen Eigenschaften des Tumors, wie zum Beispiel die erhöhte metabolische Aktivität, zum anderen liefert es Informationen zur Heterogenität im Tumor.

Im ersten Teilprojekt wurde Tumor Perfusion und Metabolismus zu folgenden Tumoreigenschaften beim HNSCC korreliert: Tumorsubarten (Oropharynx, Hypopharynx, Kehlkopf und Mundhöhle), Tumorstadium (T1/T2 versus T3/T4) und dem HPV Status. Dabei wurden Perfusions-Computertomographie (CTP) und 18F-Fludeoxyglucose Positronen-Emissions-Tomographie (18F-FDG PET) Aufnahmen von 41 HNSCC-Patienten analysiert. Drei Perfusionsmerkmale wurden ermittelt: das Blutvolumen (BV), der Blutfluss (BF) und die mittlere Durchflussgeschwindigkeit (MTT). Des Weiteren wurden Unterschiede in den Perfusionsmerkmalen zwischen dem Gross-Tumor-Volumen (GTV) und dessen Gewebeumgebung untersucht. Tumoreigenschaften, welche mit einer schlechten Prognose assoziiert sind (T3/T4 und HPV negativ), haben einen erhöhten BV-Wert sowie einen erhöhten MTT-Wert im Vergleich zum umgebenden gesunden Gewebe. Zudem wurde gezeigt, dass die Korrelation zwischen der FDG-Aufnahme und der Perfusion abhängig von den Tumoreigenschaften ist. Eine positive Korrelation wurde lediglich für den HPV-positiven ($r = 0.86$, $p = 0.04$) und oropharyngealen ($r = 0.63$, $p = 0.05$) Tumore beobachtet.

CTP Aufnahmen bestehen aus wiederholten CT-Scans und sind deshalb dosisintensiv. Wir haben eine separate Studie mit dem Alderson Phantom durchgeführt, um unser klinisches CTP-Kopf-Hals-Protokoll anzupassen. Der Endpunkt der Studie war eine Reduzierung der deponierten Dosis bei gleichbleibender Bildqualität. Unser Standardprotokoll für CTP Aufnahmen am GE Revolution CT war 100 kV, 80 mAs, 5 mm Schichtdicke und gefilterter Rückprojektionsalgorithmus. Wir haben den Prozentsatz der adaptiven statistischen iterativen Rekonstruktion (ASiR), Schichtdicke, Röhrenstrom und -spannung angepasst. Das Signal-Rausch-Verhältnis wurde in 7 vordefinierten Regionen gemessen und die effektive Dosis wurde unter Verwendung von Thermolumineszenz-Dosimetern abgeschätzt. Das optimierte Protokoll verwendet nun 80 kV, einen auf die Anatomie adaptierten Röhrenstrom von

15 bis 80 mAs, 2.5 mm Schichtdicke und 50% der ASiR-Rekonstruktion. Die effektive Dosis wurde um den Faktor 2 verringert, während die Bildqualität beibehalten wurde.

Im zweiten Teil der Doktorarbeit wurde die Radiomics Methodik verwendet, um den Behandlungserfolg und deren Korrelation zur Tumorbiologie zu untersuchen. Eine eigene Radiomics-Softwarelösung wurde in der Programmiersprache Python (Version 2.7) implementiert. In den meisten Radiomics-Studien werden Eigenentwicklungen oder Open-Source Lösungen verwendet, wobei die implementierten Berechnungsprozesse derzeit nicht vollständig standardisiert sind. Aus diesem Grund, haben wir unsere eigene Software-Lösung mit derjenigen der MAASTRO-Klinik (Maastricht, Niederlande) verglichen. mit beiden Software Lösungen haben wir ein Modell zur Prognose der lokale Tumorkontroll-Modelle erstellt. Die Radiomics Parameter wurden auf 18F-FDG PET Scans berechnet, welche 3 Monate nach Radio-Chemotherapie aufgenommen wurden (n = 128 Patienten). 649 Radiomics Feature waren in beiden Implementationen mit gleicher mathematischer Definition vorhanden, nur 80 davon konnten zwischen den Implementationen reproduziert werden (Intra-Klassen Koeffizient ICC > 0.8). In der univariaten Cox-Regressions-Analyse wurde gezeigt, dass die Aussagekraft der Radiomics Features stark von der verwendeten Implementierung abhängt. Die Hauptursachen dafür, dass viele Features nicht reproduzierbar waren können auf Unterschiede in der Maskenerzeugung, der Verwendung der Bin-Grösse in den gefilterten Bildern und dem Typ der verwendeten transformierten stationären versus nicht-stationären Wavelet-Transformation zurückgeführt werden.

In einer weiteren Radiomics Stabilitätsstudie wurde die Reproduzierbarkeit der Radiomic Features bezüglich verschiedener CTP Berechnungsfaktoren untersucht. Einige dieser Berechnungsfaktoren können nicht standardisiert werden (arterielle Eingangsfunktionsdefinition und Rausch-Level in der Berechnung) und Radiomics Parameter die nicht reproduzierbar sind unter Variation dieser Faktoren sollten nicht mit den klinischen Ergebnissen verknüpft wird. Es wurden CTP Scans von Lungen- (n = 11) und Kopf-Hals-Karzinome (n = 11) untersucht. 255 der 945 CTP Radiomic Features waren in beiden Tumorarten stabil bezüglich der Konturierung der Arterien sowie des Rausch-Levels. Nach Korrektur der Inter-Feature-Korrelation sowie der Korrelation der Features zum Tumorvolumen, konnten unter den untersuchten Features 10 Gruppen identifiziert werden. Weitere Untersuchungen werden benötigt, um die prognostische Aussagekraft jener Features zu testen.

Es wurden verschiedene Modelle gebildet um die Beziehung zwischen der lokalen Tumorkontrolle und den Radiomics Features in HNSCC zu untersuchen. Diese basierten auf CTs mit Kontrastmittel und 18F-FDG PETs, welche vor der Behandlung akquiriert wurden. Dabei wurden zwei Patienten-Kohorten verwendet: eine retrospektive Kohorte um das Modell zu bilden (n > 90 Patienten) und eine prospektive Kohorte aus einer Phase-2 Studie (durchgeführt in unserem Institut mit standardisierten Bildgebungs-Protokoll) für die Validierung des Modells (n > 50 Patienten). Tumore, welche heterogener in der CT Dichte waren, zeichneten sich durch ein erhöhtes Risiko für Tumorrezidiv aus. Dieses Modell erwies eine grössere Aussagekraft als ein Model basierend auf klinischen prognostischen Faktoren (Tumorstadium, Volumen und HPV-Status) - Konkordanzindex (CI) im der Validierungskohorte $CI_{radiomics} = 0.78$ and $CI_{clinical} = 0.73$. Auch im Vergleich zur Kombination aus CT Radiomics Parametern und klinischen Faktoren besass das Modell alleinig aus CT Radiomics eine grössere Aussagekraft ($CI_{combination} = 0.76$). In der Folgestudie, wurde untersucht, ob das Hinzufügen von metabolischen Informationen das Radiomics basierte lokale Tumorkontroll-Modell verbessert. Es wurde beobachtet, dass runde Tumore (basierend auf 18-FDG PET Autosegmentierung) mit einer fokussierten Region aus hoher FDG-Aufnahme, umgeben von einem Rand mit niedriger FDG-Aufnahme, eine bessere Prognose hatten.

Dieses Modell war jedoch nicht besser als das CT-basierte Modell. In der Validierungskohorte erreichten beide Modelle einen CI-Wert von ungefähr 0.7. Auch die Kombination von PET und CT Radiomics konnte die Vorhersagkraft nicht verbessern. Jedoch zeigte das PET-Radiomics-Modell eine bessere Kalibrierung, welche mit der Präsenz von Metallartefakten im CT im Kopf- und Halsbereich in Verbindung gebracht werden kann.

Um die abstrakten Radiomic Features mit der Tumorbiologie zu verknüpfen, haben wir die CT basierten Radiomics Parameter mit dem HPV-Status korreliert. Dabei wurde beobachtet, dass HPV positive Tumore eine homogenere CT -Dichte aufweisen. Obwohl das Set von Radiomic Features das den HPV Status beschreibt eine ähnliche Interpretation zu den Parametern, die lokalen Tumorkontrolle charakterisieren, aufweist, umfasst sie verschiedene Merkmale und die Parameter waren nicht miteinander korreliert. Zum Beispiel war das lokale Tumorkontroll-CT-Radiomics Modell auch in der HPV-negativen Untergruppe von Patienten prognostisch.

Zusammenfassend haben wir gezeigt, dass biologische Informationen auch aus der einfachen morphologischen Bildgebung (CT) wiederhergestellt werden können. Zusätzlich haben wir bildgebenden Eigenschaften basierend auf Perfusionsunterschieden zwischen dem Tumor und seiner Umgebung sowie CT- und PET-basierenden Radiomics identifiziert, die mit einer schlechteren Prognose assoziiert waren. Diese Signaturen sollten in einer externen Kohorte von Patienten weiter validiert werden. Zudem müssen Optionen zur Intensivierung der Behandlung für schlechtere Prognosegruppen definiert werden.

Table of context

Chapter 1. Introduction

1.1 Cancer

1.2 Radiotherapy

1.3 Medical imaging in radiotherapy

1.5 Head and neck cancer

1.6 Radiomics

Chapter 3. Results

Chapter 4. Discussion and outlook

Chapter 1

Introduction

1.1 Cancer

1.1.1 Epidemiology of cancer

Almost 4 million new patients are diagnosed with cancer each year in Europe [5]. Yearly it causes around 20% of deaths [5]. Globally, the number of new cases is expected to rise by about 70% over the next 2 decades [6]. The rate of 5-year overall survival greatly varies between tumor sites, stages and subtypes [3, 4, 7]. It ranges from less than 10% in pancreatic cancer to more than 90% in prostate cancer in the USA [7].

Three factors can lead to induction of mutations and consequently carcinogenesis: physical, chemical and biological factors [5]. Ionizing radiation is an example of a physical factor. Chemical factors comprise: asbestos, tobacco smoke, aflatoxin and arsenic. Biological factors are connected to infections from viruses, bacteria or parasites. Therefore, the common cancer risk factors include: smoking, unhealthy diet, alcohol abuse, HPV-infection, infection by hepatitis, overweight, lack of physical activity, exposure to ionizing and ultraviolet radiation as well as urban air pollution.

1.1.2 Hallmarks of cancer

Cells in a multicellular organism function as a society and are committed to collaborate with each other [8]. They send and receive signals, which coordinate their resting, growing, dividing, differentiating and eventually dying. Mutations can influence those control mechanisms. Carcinogenesis is a multistep process, when a cell accumulates mutations, which enable an increase of division rate and survival in a hostile environment. This selective advantage promotes survival of that cell at the expense of its neighbors. All cancer cells share common traits (hallmarks) that enable tumor growth and metastatic dissemination [9]. They are summarized in the Figure 1. They can be divided into eight biological capabilities and two enabling characteristics.

Cancer cells can both sustain proliferative signaling and evade growth suppressors [9]. Unlike the normal cells, cancer cells do not follow a normal regulatory mechanism where cells control the production and release of growth-promoting signals. Therefore, they escape the maintenance of cell number and tissue architecture. Additionally, they are blocking pathways preventing cells from proliferating. Moreover, cancer cells have the capability to resist cell death. They are resistant to extracellular and intracellular signals inducing apoptosis. Normal cells are programmed to a limited number of divisions. However, cancer cells present replicative immortality. The rapid proliferation and high cellular concentration would not be possible without a supply in oxygen and nutrition. Therefore, cancer cells need to develop the capability to form their own vasculature and to induce angiogenesis. This process is rarely switched on in the normal tissue (for example during a wound healing), but is almost constantly on in tumors. The high nutrition supply is essential for the reprogramed energy metabolism. Cancer cells were shown to metabolize glucose in the glycolic process even in the presence of oxygen, so called aerobic glycolysis. The aerobic glycolysis is known to be less efficient but a faster process than conversion to pyruvate and carbon dioxide. Cancer cells can recruit more glucose and compensate for the aerobic glycolysis inefficiency [10]. A rapidly growing tumor needs more nutrition and more space. Cancer cells are able to alter their shape as well as the attachment to other cells and extracellular matrix via the epithelial-mesenchymal transition, which enables them an active invasion and development of metastasis [9]. The last biological capability of cancer cells is the avoiding of immune destruction.

The two enabling characteristics support the acquiry of aforementioned distinct biological capabilities [9]. A transition from normal to cancerous cell requires multiple steps and multiple mutations. Cells, which inherits an impaired DNA repair mechanism or resistance to cell death are more prone to genomic instability and further mutations. Additionally, the inflammatory state in the tumor microenvironment enhances the infiltration, carcinogenesis and progression, by for example suppling the tumor microenvironment in growth or proangiogenic factors.

Although common cancer traits exist, they can be caused by a wide range of mutations and consequently by alteration of different regulatory pathways. It results in a high intra- and inter-tumor heterogeneity.

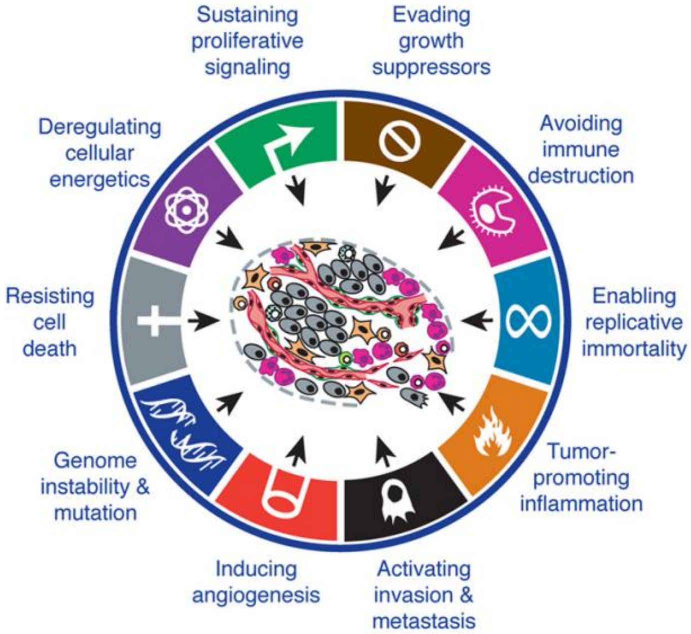


Figure 1. The summary of hallmarks of cancer adapted from [9].

1.1.3 Heterogeneity of cancer

Cancer is a heterogeneous disease, showing intra- and inter-patient genetic and phenotypic variation. Tumor cells show a differential gene expression due to genetic and epigenetic heterogeneity [1]. It can further correspond to differences in the cell proliferation rate, tumor aggressiveness and its level of hypoxia. Currently, a comprehensive explanation for translation of the genomic instability to intra-tumor heterogeneity and sub-clonal architecture is lacking. In addition, the fluctuations in tumor microenvironment and exposure to a treatment may contribute to intra-tumor heterogeneity. Tumor biopsy, functional imaging and blood or other body fluids analysis can quantify the biochemical or molecular genetic substances (biomarkers) corresponding to disease phenotype [11]. Different phenotypes may lead to differential response to a treatment and thus in-depth understanding of cancer heterogeneity is crucial for the development of new treatment strategies as well as the choice of the right treatment.

1.2 Radiotherapy

1.2.1 Principle

Radiotherapy is one of the main treatment options in cancer, next to surgery, chemotherapy, targeted therapy and immunotherapy. The aim of radiotherapy is to deliver the prescribed dose of ionizing radiation (energy deposited per mass) to a target volume (for example tumor) with minimal damage to surrounding healthy tissues. To achieve that, it is crucial to understand how radiation interacts with matter and how it influences function of a living organism.

Ionizing radiation is any type of radiation that carries energy high enough to liberate electrons from the atoms and thus create ions. The most common type of ionizing radiation used in radiotherapy are photons. Photons transfer their energy to atoms in an absorber, which leads to the creation of a secondary radiation and further ionization of the material. There are three main types of photon interactions with matter: photoelectric effect, Compton scattering and pair production. Their probability (cross section) depends on the photon's energy and composition of the medium. In the photoelectric effect, a photon transfers its energy to an electron bound in an atom. The part of the energy is used for breaking a bond between the electron and the atom, and the rest is transformed to the electron's kinetic energy. Compton scattering is an inelastic scattering of a photon on a free electron. The photon transfers a part of its energy to the electron, which results in change of the photon's travel direction. Pair production occurs for photons with energy higher than 1.02 MeV. In the presence of a third body (electron or nucleus) a photon can disappear and reappear as a pair of electron and positron. The particles are emitted in the opposite directions and energy of each of the particles equals half the energy of the photon minus the electron rest mass.

Most of the patients are treated using linear accelerators. In the linear accelerator electrons are accelerated to energies from 4 to 18 MeV. To produce photons a target made from a material with high atomic number is placed on the path of electrons [12]. In the interaction of electrons with target photons are produced via bremsstrahlung. Photon energy spectrum from a linear accelerator is heterogeneous and its maximal energy is equal to maximal energy of the used electron beam. The Compton effect is a dominant interaction in this photon energy range and soft tissue medium (Figure 2). Free electrons produced in the photon interactions ionize other molecules. At the end of their path, they create clusters of ionizations. If they occur in the region of DNA they are potentially lethal.

On a time scale, three phases of interaction of ionizing radiation with a cell are distinguished [13]. First, ions (free radicals) are produced in the physical phase. The stable equilibrium is lost. Free radicals are highly reactive. They interact with other molecules disrupting their structures e.g. create single- and double-strand breaks in DNA (chemical phase). The second phase lasts few seconds. The third, biological, phase is the longest and ranges from seconds to years. It includes all the reactions of the organism on the damage, from repair processes, early and late side effects, to possible secondary carcinogenesis. This last phase is the field of studies in radiobiology.

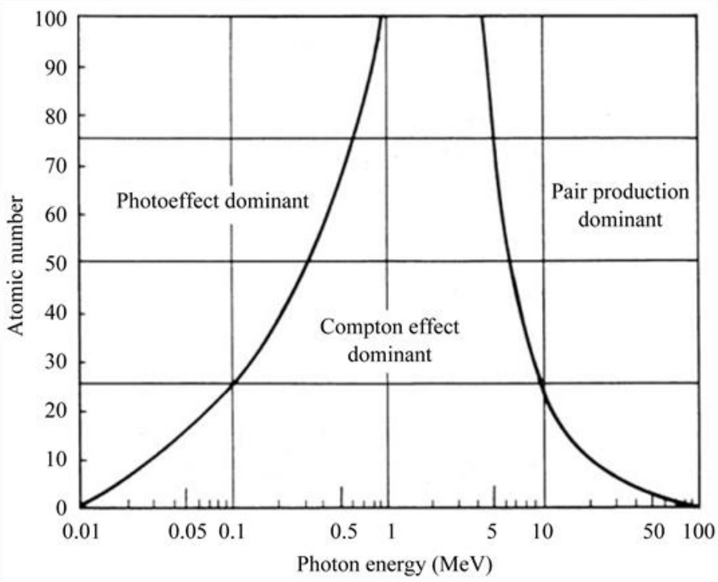


Figure 2. Regions of relative predominance of the three main forms of photon interaction with matter. The Compton scattering is predominant interaction in water for photons in the therapeutic energy range [12].

1.2.2 Radiobiology

Radiobiology studies response mechanisms of a living organism to the damage caused by ionizing radiation. Not every irradiated cell will die. On average, the dose of 1Gy will kill only around 30% of the cells [13]. The 5Rs of radiobiology (repair, redistribution, reoxygenation, repopulation, and intrinsic radiosensitivity) summarize factors influencing cytotoxicity of ionization radiation.

Cells have the capacity to repair DNA damage and thus not every ionization in the DNA base region is lethal. The examples of DNA repair mechanisms in response to irradiation are: base excision repair, single-strand break repair, non-homologous end joining, and homologous recombination [8]. The majority of breaks are repaired within the first few hours after the irradiation. A sensor group of proteins searches through the genome for a damage and signal it to three main effector pathways: checkpoint, repair and death. Some of the checkpoints may be disabled in a cancer cell, causing cell replication in the presence of DNA damage. Consequently, it may increase genomic instability [13]. Repair mechanism depends on the cell cycle phase. For example, homologous recombination uses the sister DNA strand as a template for repair. The sister DNA strand is available in S and G2 phases and cells in these cycle phases are known to be more resistant to radiation damage.

A consequence of differential repair efficacy through the cell cycle is redistribution. As the cells in the S phase are the least radiosensitive, the majority of surviving cells were in the S phase in the moment of irradiation. This causes a synchronization in the cell cycle of the tumor cells. However, it is lost after some time and thus cannot be exploited as a target for radiosensitization (for example irradiation at the time points when most of the cells are in the most radiosensitive cell cycle phase). It seems that the cancer cells proceed through the cell cycle with heterogeneous speed.

Oxygen concentration in the cell also influences the radiosensitivity. Free radicals ($R^\bullet$) produced by the ionizations are highly reactive and contributes greatly to the DNA damage caused by irradiation. In the presence of oxygen they produce $RO_2^\bullet$, which ultimately results in a stable change in a target (ROOH).

In the absence of oxygen, they have a longer half-life and can react with H^+ ions to restore its original form. In this case they do not contribute to DNA damage. In consequence, the well-oxygenated cells are more radiosensitive and thus are primarily killed. It causes the reoxygenation of hypoxic cells as they have a better access to blood supply. Hypoxia-modifying agents are promising tools for radiosensitization [14].

Decrease in cancer cell numbers also has a negative effect, as it induces repopulation [13]. It can be related to increased access to nutrition or a regenerative response. The doubling time shortly after irradiation exceeds doubling time of non-irradiated cells. A rapid repopulation is observed after around 3 weeks of fractionated therapy, which can be an effect of increased proliferation or decreased cell loss. As a consequence of repopulation, higher cumulative dose has to be delivered to kill the tumor.

Evidence from studies in cell lines, animal tumor models and clinical samples shows a wide variation in intrinsic tumor radiosensitivity. Tumors of similar origin and histological subtype show a differential response to irradiation. The intrinsic tumor radiosensitivity is defined as an in-vitro cell survival fraction of cells from individual tumor after 2Gy irradiation. It is was shown to correlate with treatment outcome [15]. In the correlation with gene expression it was linked to genes responsible for regulation of DNA damage response, histone deacetylation, cell-cycle, apoptosis and proliferation.

The 5R’s of radiobiology points out the tumor characteristics, which are linked to higher radioresistance. Targeting these characteristics could help to improve treatment outcome

1.3 Medical imaging in radiotherapy

1.3.1 *Basics of morphological, functional and molecular imaging*

Medical imaging plays an important role in the management of cancer. It has a wide range of applications from screening and diagnosis to tailoring treatment decision [16]. In radiation oncology each treatment plan is based on the individual patient anatomy and tumor localization, which is assessed by medical imaging. Three types of imaging can be distinguished: morphological, functional and molecular [17]. This section gives a short introduction to the basics of medical imaging techniques studied in this PhD project: computed tomography, computed tomography perfusion and positron emission tomography.

Computed tomography

The common and most often used example of morphological imaging is computed tomography (CT) [18]. It is based on the analysis of X-ray attenuation in different tissues. The energy range used in CT imaging is in order of 100 kV. The photoelectric effect is the dominant interaction in this energy range. Its mass attenuation coefficient is proportional to atomic number (Z) to the power of 3, therefore the photon absorption is higher in tissues with high Z. The Hounsfield scale relates a tissue linear attenuation coefficient (μ), the direct measure in CT, to linear attenuation coefficients of water (μ_{water}) and air (μ_{air}):

$$HU = 1000\frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \tag{1}$$

It gives a relative measure of number of photos passing through a tissue, which is also called radiodensity. This property is used in the radiotherapy planning process to calculate delivered dose.

Computed tomography perfusion

A repeated contrast-enhanced CT scan can also provide functional information. Capillary level hemodynamic information, mainly blood flow (BF), blood volume (BV) and mean transit time (MTT), can be derived from repeated CT scans of the same region of interest, which shows the time-dependent distribution of intravascular iodine contrast agent [19]. A theoretical physiological model of tissue perfusion is used to estimate the perfusion parameters [20]:

$$C_v(t) = \int_0^t C_a(\tau)R(t-\tau)d\tau \tag{2}$$

where: C_v – concentration of contrast agent in the voxel, C_a – concentration of the contrast agent in the supplying artery, R – residue function describing the amount of contrast still present in the voxel at time τ. Knowing the time-dependent concentration in the voxel and in the supplying artery the deconvolution can be performed to solve the equation 3 for *R(t).* The perfusion descriptors are consecutively defined as:

$$BF = \max\big(R(t)\big) \tag{3}$$

$$MTT = \frac{1}{max\big(R(t)\big)} \int_0^\infty R(t)dt \tag{4}$$

$$BV = BF \cdot MTT \tag{5}$$

The equation (5) is a consequence of the central volume theorem [21].

Positron emission tomography

The positron emission tomography (PET) is an example of molecular imaging, where the observed process depends on a compound administered to the patient (radiopharmaceutical) [22]. The radiopharmaceutical (tracer) consists of a β^+ radioisotope and a ligand – biochemical substance of a receptor system. The positron emitted in β^+ decay travels on a short distance until it loses the kinetic energy and meets an electron, which results in an annihilation process. Two photons of energy of 511 keV are emitted in opposite directions. The coincident detection of those photons by a detector ring is used to define the emission point and thus the distribution of the radiopharmaceutical in the body. The example of PET radiopharmaceutical widely used in oncology is 18F-fludeoxyglucose (18F-FDG) the analog of glucose labeled with radioactive 18F. It helps to identify regions of increased metabolism in the body. The activity measured by the PET scanner depends on patient size and the injected 18F-FDG activity. The standardized uptake value (SUV) was defined to account for these factors and increase the comparability between the measurements [23]:

$$SUV = \frac{A_v}{A_i} w \tag{5}$$

where: A_v – measured activity in a voxel, w – body weight, A_i – injected activity corrected for decays during the uptake time.

1.3.2 Staging and target definition

The well-established staging system in cancer - TNM (tumor-node-metastasis) scores the primary tumor size, its invasion of the adjacent structures, the involvement involved lymph nodes and their distribution as well as the presence and location of distant metastasis [24]. Cancer diagnosis requires the pathological examination, whereas imaging is crucial for disease extent assessment. The right choice of treatment depends on the correct TNM classification. The addition of functional imaging (18F-FDG PET/CT) to anatomical imaging improved TNM staging and consecutively treatment outcome

[24, 25]. The 18F-FDG-PET imaging is especially useful in detection of suspicious lymph nodes and asymptomatic distant metastasis.

The introduction of CT imaging in 1980s triggered a change in target definition in radiotherapy [22]. In 1993, the International Commission of Radiation Units and Measurements released a report with definition of volumes, which need to be considered in the external radiotherapy [26]:

- the Gross Tumor Volume (GTV) - the gross palpable or visible/demonstrable extent and location of the malignant growth. The demonstrable extent can be visualized using different medical imaging techniques (e.g. computed tomography, magnetic resonance, positron emission tomography, mammography and ultrasound).
- the Clinical Target Volume (CTV) contains the GTV and subclinical microscopic malignant disease.
- the Planning Target Volume (PTV) is created to compensate for uncertainties of the tumor position caused by patient set-up errors, patient movement during treatment fraction or due to internal organs movements. In the 1999 revision, the PTV was separated into two components: the Internal Target Volume (ITV) as a margin around CTV to account for variation in CTV size and location (due to the motion of adjacent organs); and PTV as a margin to account for set-up uncertainties [27].
- the Treated Volume is defined by the isodose corresponding to the dose appropriate to achieve the tumor eradication.
- the Irradiated Volume is the volume, which receives a significant dose in relation to normal tissue tolerance.
- organs at risk are all healthy tissues which would affect the treatment planning process.

Over the years, development in medical imaging techniques allowed for more precise target and organs at risk definition. Together with technical developments in radiotherapy, such as the introduction of the intensity modulated treatment, it resulted with a higher personalization of treatment plans, dose intensification and lower rate of side effects [22]. The exact definition of tumor borders is challenging and often depends on the experience of radiation oncologist. The use of additional functional information from 18F-FDG PET was shown to reduce the interobserver variability in different tumor types [28-30]. In the case of tumors originating from organs with high natural concertation of glucose, other PET traces are available, for example amino acids for brain and choline for prostate [22]. Additionally, the molecular imaging can be used to define radioresistant subregions of tumor (concept of biological target volume [31]), such as regions with increased metabolism or hypoxia, to deliver a higher radiation dose in a dose painting approach [32-34]. Precise tumor localization may also be challenging due to its inter- and intra-fraction motion caused by differential organ filling or respiratory motion. The introduction of cone-beam CT and 4D-CT helped to reduce the associated uncertainty.

1.3.3 Response assessment

Tumor response is an important endpoint in further disease management as well as the design of prospective clinical trials [35]. The reduction of tumor burden can be one of the indicator of tumor response. Response Evaluation Criteria in Solid Tumors (RECIST) defines the methodology for measurements of tumor shrinkage or progression. It points out medical imaging (computed tomography, magnetic resonance and ultrasound) as an important component of that process. The follow-up FDG-PET imaging shows a high sensitivity in detection of recurrences [36]. However, false positive findings can be observed due to the post-radiotherapy inflammatory processes [37].

The 3'-deoxy-3'[(18)F]-fluorothymidine (FLT), marker for cell proliferation, presented a good differentiation between post-treatment necrosis and tumor recurrence in glioblastoma [38].

Imaging can also be performed in the interim of treatment to assess if further intensification is needed. A prospective phase II study showed that the continuation of preoperative chemotherapy in esophageal cancer for patients, who did not show early 18F-FDG uptake decrease was not beneficial [39]. However, definition of pathologic complete response using imaging after neoadjuvant treatment, to avoid unnecessary surgery, remains challenging [40, 41].

1.3.4 Imaging tumor biology

Functional imaging serves as a 3D surrogate of tumor biology and this contributes to its wide applicability in radiation oncology, from target definition and treatment adaption to response assessment. The two tumor characteristics studied in the largest extent are metabolism and oxygenation.

The 18F-FDG PET is a widely used imagining modality for tumor metabolism quantification [22]. However, a high natural glucose accumulation in some organs triggered a research in the field of tracers for tumor-specific metabolism unrelated to glycolysis. The 11C-Acetate is linked to metabolism and has no urinary excretion, thus can be used for imaging of prostate and other urological diseases. The 11C-Choline or amino acids analogs (for example 18F-DOPA) enable visualization of brain tumors. Not only PET tracers can be used to detect tumor metabolites. The magnetic resonance spectroscopy can track similar processes as described above. A detection of substances such as choline and lactate (product of glycolysis) is possible.

Several PET tracers related to hypoxia are currently available in the clinics. The 18-fluoromisonidazole (18F-FMISO) is a widely studied tracer, which was validated for different tumor sites and its correlation with regional partial oxygen pressure was confirmed [42]. Further, a newer generation tracers 18F-FAZA and 18F-HX4 are promising hypoxia markers with a superior pharmacokinetics [22, 43]. In the field of magnetic resonance (MR) imaging, it was observed that deoxygenated hemoglobin has paramagnetic properties and thus causes magnetic field inhomogeneity. This effect can be measured by a blood oxygen level dependent (BOLD) MR sequence. However, in most cases the results are only qualitative and quantitative comparison is impossible [22]. The CTP allows for a quantitative measurement and blood volume was shown to correlate with microvessel density in colorectal cancer [44].

Tumor biology influences its response to radiotherapy. Pretreatment imaging can serve as a surrogate of tumor phenotype. Therefore, a large effort is made to predict treatment outcome based on the pretreatment imaging characteristics. Few examples are mentioned in **Chapter 1.5.2**.

1.4 Precision medicine

Precision medicine aims to tailor treatment to intra-disease specific traits. An old example of precision medicine is blood transfusion, which has been guided by blood types since decades [45]. Precision medicine in cancer is a rapidly growing field of research with genomics, proteomics, metabolomics and studies of cellular assays being major drivers in understanding the complexity of the disease. Additionally, development of advanced computational tools is crucial to extract the relevant information from large databases. Targeted therapies acting on a specific molecular target are under development [46, 47]. It also changes the design of clinical trials, from large, population-based studies to prove treatment efficacy to a targeted approach [48].

The choice of the right treatment depends largely on the development of biomarkers, measurable indicators of some biological state or condition [49]. Two types of biomarkers can be distinguished: prognostic and predictive [50]. Prognostic biomarker gives information about overall risk of recurrence of disease after a treatment, whereas the predictive biomarker indicates a benefit of a given treatment for a subgroup of patients. For example, prostate-specific antigen level correlates with patient response to prostate cancer therapy, thus it is a prognostic biomarker [51]. On the other hand, treatment with Trastuzumab provides benefit only for HER2 overexpressing breast and gastric/gastroesophageal cancer, thus HER2 expression is a predictive biomarker [52].

Cancer is a dynamic disease, which exhibits clonal evolution, intra- and inter-tumor heterogeneity as well as varied mechanisms of drug resistance. Therefore, an ideal cancer biomarker should allow for spatial tumor analysis and be non-invasive to assure repeated tumor monitoring. This is a major challenge in cancer biomarker development. Promising candidates are the imaging-based biomarkers [53].

1.5 Head and neck cancer

1.5.1 Overview

Head and neck cancer is originating from: lip, oral cavity, nasopharynx, oropharynx, hypopharynx, larynx or trachea. Head and neck cancer accounts for around 5-10% of new cancer cases in developed countries [2]. The 5-year overall survival rate depends on the tumor site. It is more than 90% in the case of lip cancer, but less than 40% in hypopharyngeal carcinoma [54]. However, over the last few decades the prognosis for most of the head and neck caner subtypes improved. The introduction of concurrent radiochemotherapy was shown to improve overall survival in comparison to radiotherapy alone [55]. Moreover, in recent years the incidence of human papillomavirus (HPV) related tumors, showing a better response to a therapy, is increasing [54].

The major risk factors include: tobacco use (also chewing tobacco), alcohol consumption, HPV infection, male gender and betel quid chewing [56]. The combination of alcohol and tobacco abuse was observed to have a large impact than alcohol only. The HPV infection, as a cause of head and neck cancer, was mainly linked to the oropharyngeal carcinoma. The geographical variation in the incidence of head and neck cancer depends on the habits of tobacco and alcohol abuse, which contributes to the development of around 80% of new cases diagnosed globally. In recent years, head and neck cancer rate has been decreasing for the alcohol consumption related cases, but has been increasing for the HPV-related ones [56].

The recommended treatment depends on the primary tumor location and extension [57]. In early stage tumors (I-II) a similar rate of loco-regional control can be achieved by conservative surgery or radiotherapy. In the locally advanced stage III and IV tumors surgery combined with postoperative radiotherapy or chemoradiotherapy is an option. In the case of non-resectable tumors or patients with contradiction for surgical treatment, combined concomitant chemoradiotherapy is offered. This treatment option is also offered in resectable patients with an anticipated poor functional outcome. In comparison to surgery it is an organ preservation treatment option with a similar treatment outcome [58]. Although the introduction of advanced radiotherapy techniques, such as intensity modulated radiotherapy, increased treatment conformity, the toxicity rate is still high [59, 60]. The common toxicities comprise: mucositis, dermasitis, dysphagia and xerostomia. Several clinical studies are currently investigating different treatment protocols to decrease the toxicity and increase treatment efficacy [61]. The exemplary concepts include: adaptive radiotherapy, altered fractionations, reduction of high dose target volume (dose painting) for selected patients, or de-escalation of dose for HPV positive oropharyngeal carcinoma.

1.5.2 Prognostic factors and imaging biomarkers in head and neck cancer

Head and neck cancer squamous cell carcinoma (HNSCC) shows a heterogeneous response to radiochemotherapy with loco-regional control and 5 years overall survival ranging from below 50% to 80% [3, 4]. The human papilloma virus (HPV) is a well-established prognostic factor in the subgroup of oropharyngeal carcinoma. The HPV infection corresponds to a favorable radiotherapy treatment outcome [3, 4]. The overexpression of epidermal growth factor receptor (EGFR) generally relates to poor prognosis, however those tumors show a better response to hyperfractionated and accelerated radiotherapy in comparison to a standard one [62]. Tumor hypoxia (largely mediated by HIF-1) makes it more resistant to radiotherapy [63]. In terms of chemotherapy, the overexpression of cyclin D1 was shown to cause cisplatin resistance.

Several groups have investigated pretreatment medical imaging as a biomarker in head and neck cancer. Mixed results have been reported in respect to prognostic power of 18F-FDG maximum SUV [64]. However, other 18F-FDG measures: metabolic tumor volume (MTV) and tumor lesion glycolysis, were found to be prognostic in meta-analysis for adverse events and overall survival [65]. Few studies have shown that, the uptake of PET hypoxia tracers (18F-FMISO and 18F-FAZA) correlates with disease free survival [64]. Further, tumors with higher cell density, quantified by increased apparent diffusion coefficient in MR, were linked to worse prognosis [66]. Recently, a single center study investigated multiparametric imaging 18F-FDG and 18F-FMISO. It showed that the slope of voxelwise correlation between the uptakes of those two tracers is prognostic for tumor recurrence [67].

Medical imaging enables to capture the 3D complexity of a tumor. Recent studies try to correlate tumor heterogeneity with treatment outcome. Those efforts are described **in Chapter 1.6**.

1.6 Radiomics

1.6.1 Principle of radiomics

Radiomics aims for a quantitative and comprehensive description of a medical image [17]. The quantitative analysis is assured by the use of well-described mathematical definitions to extract features such as contrast or heterogeneity [68]. In comparison to traditional image assessment, radiomics is observer independent. The comprehensive description is achieved by the extraction of hundreds of features from the region of interest (ROI). This high-throughput analysis allows for studying differences between the ROIs on different scales and using different characteristics. Four main radiomic features characteristics are distinguished: shape-, intensity-, texture- and filter-based features (Figure 3).

The shape-based features describe the properties of the ROI outline. Among others, they provide information about the ROI's volume, surface, sphericity and compactness [69]. They do not contain any information about image intensities.

The intensity-based features (also known as first-order texture features) provides global description of image intensities in the ROI. They comprised well-known histogram descriptors such as: mean value, minimum value, maximum value, skewness and kurtosis [69]. They ignore the relationship between the adjacent voxels.

The texture-based features focus on the intensity changes between the adjacent voxels. Examples of texture features are contrast, homogeneity, directionality and coarseness [69]. Texture features are calculated in two steps. First, a matrix describing dependencies between the voxels is defined. Depending on the number of voxels taken into account in the calculation we distinguish second-order texture features (two voxels) and higher-order texture features (three and more voxels). An example of second-order matrix is the Gray Level Co-occurrence Matrix (GLCM), which describes intensity changes between two voxels in a defined distance and direction [70]. One way to calculate the higher-order texture features is to compute the Gray Level Size Zone Matrix, which contains information about the number of homogenous regions of a given size and intensity level [71]. In the second step, these matrices are used to compute the texture characteristics.

The filter-based features allow for analysis of the image in different scales and in the frequency domain. The most popular filters are wavelets and Laplacian of Gaussian (LoG) [72, 73]. The wavelet transform enhances the information about low and high frequencies in the image. The LoG filter enables detection of edges. After the image filtering, the intensity- and texture-based features are computed on the new images. Altogether, filter-based features allow for in-depth analysis of image heterogeneity.

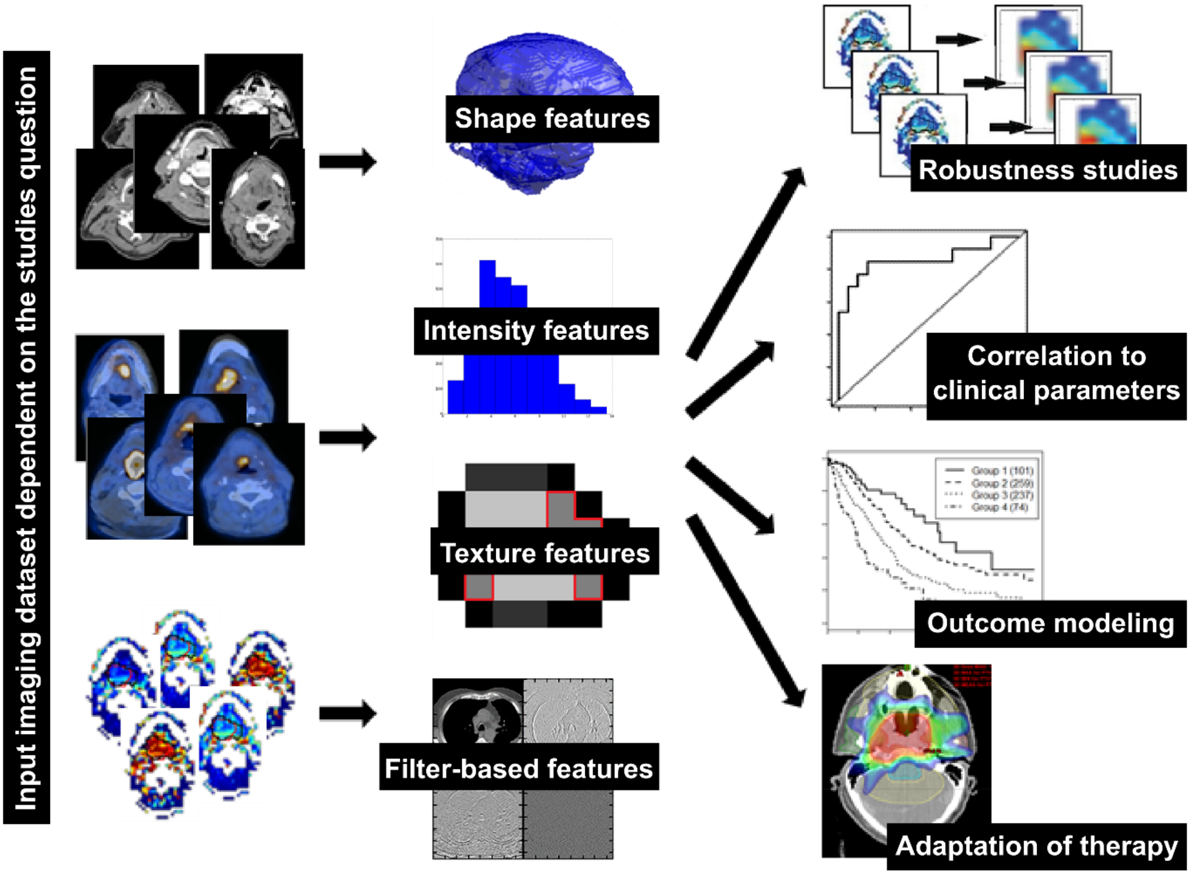


Figure 3. Scheme of radiomics analysis. Depending on the studied question, different input data (image modalities) can be used.

1.6.2 Radiomics prognostic signatures

The potential of radiomics to predict treatment outcome has been recognized for many tumor types and using different imaging modalities [69, 74-85]. In addition, multiple endpoints were studied, ranging from loco-regional control to development of new metastases. One of the first, large radiomics studies has shown that CT radiomics is prognostic for patient overall survival in lung cancer after curative radiotherapy or chemoradiotherapy [69]. The same signature was further validated for head and neck cancer and its prognostic power was also proven in the presence of CT artifacts [69, 79]. Although there seems to be a general radiological phenotype associated with worse prognosis, more recent study showed that some of the radiomic features are prognostic only for certain tumor types [80].

Clinical relevance of radiomics is under investigation. Radiomic features were found to significantly improve models based on clinical parameters. The addition of previously mentioned radiomic signature to the tumor volume and the TNM stage increased the model prognostic power in both lung and head and neck cancer [69]. More recently it was shown, that the inclusion of another CT radiomic signature to a clinicopathological nomogram (tumor stage, histologic grade, gender, age) improves prediction of disease free survival in patients with early stage lung cancer after surgical resection [76]. Another example for clinical relevance of radiomics is the identification of new non-invasive biomarkers for certain treatment options. In recurrent glioblastoma, a group of patients who may potentially gain the most benefit from antiangiogenic therapy was identified using MR radiomics [77, 85]. In the locally advanced rectal cancer, the MR defined tumor compactness allowed for better identification of patients with and without complete pathological response after the preoperative concurrent chemoradiotherapy [86]. Similarly, in breast cancer, FDG-PET radiomics was associated with pathological complete response after neoadjuvant chemotherapy [75].

1.6.3 Radiomics link to tumor biology

The complex nature of radiomic signatures and the fact that radiomic features often outreach human visual perception is a motivation for studies relating radiomics to tumor biology. The histopathology or genetic profiles are used to better interpret radiomic signatures. In lung cancer, CT radiomic features corresponding to tumor heterogeneity were found to correlate with cell cycling pathways [69]. Further, the influence of two important biomarkers: epidermal growth factor receptor gene (EGFR) and KRAS protein, on radiomic phenotype was investigated. Both PET and CT radiomics were shown to successfully distinguish the $EGFR^+$ and $EGFR^-$ lung tumors [87, 88]. On the other hand, the KRAS status did not influence FDG-PET tumor phenotype and the changes in CT phenotype were found to be moderate.

Similar studies were performed for other tumor types. In glioblastoma, both IDH1 and EGFR status were shown to influence the tumor MR phenotype [89, 90]. The metabolic patterns, quantified by 18F-FDG-PET radiomics, in breast cancer were correlated with the Ki67 expression, a molecular biomarker for cell proliferation [75].

The correlation of radiomics with tumor biology may be challenging as it requires a large number of both imaging and molecular data. An analysis of preclinical data could be a solution to that problem. Significant changes in CT phenotype were observed in tumor xenografts with differential hypoxic fraction [91].

1.6.4 Radiomics challenges

Radiomics is a relatively new discipline. The major challenges in radiomics come from missing standardization of: image acquisition parameters, definitions of radiomic features, normalization methods and reporting of results and methods. Radiomic features need to be validated against image acquisition and post-processing protocols [92-95]. Results of some of the stability studies, such as voxel size influence on texture, can be translated between image modalities. Nevertheless, most of the factors are modality-specific and warrant separate investigation. It was recently shown that even the best method for image discretization is modality dependent [92]. Additionally, little is known about the translation of stability results from one tumor site to another.

Several radiomics implementations are currently available. Most of them are an open source or in-house developed codes. Although they are based on the same mathematical definitions, the freedom in their implementation is a big challenge in the field [93, 96]. Studies aiming for workflow standardization should be encouraged [68].

The challenges in radiomics in the context of robustness studies performed within this PhD project are further addressed in **Chapter 4.3**.

Chapter 2

Aims

This PhD project aims to investigate imaging phenotypes in head and neck squamous cell carcinoma (HNSCC) related to outcome prognosis in radiotherapy. It consist of two main parts: the investigation of the influence of tumor subtype on functional imaging phenotype and the study of imaging phenotypes related to prognosis and tumor biology by means of radiomics.

In the first part, we hypothesize that HNSCC tumor imaging phenotype, in particular the metabolism and perfusion, depends on tumor subtype, stage and HPV status **(Chapter 3.1)**. Previous studies found that, the HPV infection interfere with the tumor microenvironment, for example HPV positive tumors show more rapid decrease in their hypoxic fraction after irradiation [97]. Additionally, head and neck cancers grow from differentially vascularized tissues, which may subsequently influence vascularization of a tumor [98]. Changes in the vasculature may interfere with glucose transport and glucose metabolism. I have assessed tumor metabolism using 18F-FDG PET and perfusion using CT perfusion (CTP). I have studied the differences in perfusion parameters between tumor and the surrounding healthy tissue, the differences in perfusion and metabolism for HNSCC subgroups as well as correlation of perfusion and metabolism for the HNSCC subgroups. The CTP is a dose intense examination. Therefore, I have further investigated dose reduction methods (iterative reconstruction and tube current and voltage adaptation) for a head and neck CTP protocol **(Chapter 3.2)**. This subproject aimed to define a new CTP protocol with a similar image quality, improved resolution and lowered dose.

In the second part of the thesis, I have exploited radiomics (advanced image analysis technique) to capture comprehensive tumor imaging subtypes. This PhD project contributes to the field of radiomics by evaluation of its robustness as well as by its correlation to treatment outcome and tumor biology.

The first two studies are focused on the radiomics standardization and robustness. Majority of radiomics studies are conducted using open source or in-house developed implementations. A standard guideline for radiomics implementation is not yet defined (detailed description of my implementation can be found in **Chapter 3.3** and **Chapter 4.2**). Therefore, I have investigated the influence of independent radiomics implementation on the prognostic power of radiomic features **(Chapter 3.3)**. The two studied implementations were based on the same mathematical definitions. I have compared the local tumor control predictions between the implementations as well as tried to identify differences in the radiomics workflow that contribute to the errors in those predictions. The other source of errors in radiomics-based predictions is image quality. Image quality in CTP can be influenced by multiple factors as it is a computationally complex technique. Some of the computation steps are difficult to standardize, for example artery contouring or noise threshold. Therefore, the second study was focused on the stability of radiomics features in CTP maps. We hypothesize that CTP computation steps influence the stability of radiomic features thus reducing the set of the features, which can be used for prognostic modeling **(Chapter 3.4)**. I have evaluated this hypothesis for lung and head and neck perfusion scans.

Finally, I have performed two studies to link radiomic signatures (set of radiomic features) with tumor biology and local tumor control prediction to provide better understanding of imaging phenotypes. Some exploratory studies have already indicated a relationship between HPV infection and the heterogeneity of imaging-based tumor density [99, 100]. However, they were performed on small cohorts of patients with no validation. In this project, I have used CT data from 149 patients divided into two separate cohorts to investigate the link between HPV status and CT radiomics **(Chapter 3.5)**. In the outcome modeling, most of the radiomics-based studies uses the overall survival as an endpoint [101]. Whilst it is important from the clinical point of view, it is also influenced by other non-tumor-related factors [102]. I have therefore decided to study local tumor control as our endpoint. First, I have investigated the added value of a CT radiomics signature for local tumor control to the clinically relevant prognostic factors: tumor volume, tumor stage and HPV status **(Chapter 3.5).** In the second step, I have evaluated whether tumor metabolic information improves tumor control predictions in comparison to CT radiomics **(Chapter 3.6)**. Conflicting results have been published on the role of pretreatment 18F-FDG PET (mean and maximum FDG uptake) for prediction of treatment outcome in head and neck cancer [103]. I have therefore quantified tumor metabolism using PET radiomics and subsequently correlated it with treatment outcome and compared it with predictions based on CT radiomics.

Chapter 3

Results

3.1 Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma.

Nesteruk Marta[1], Lang Stephanie[1], Veit-Haibach Patrick[2,3], Studer Gabriela[1], Stieb Sonja[1], Glatz Stefan[1], Hemmatazad Hossein[1], Ikenberg Kristian[5], Huber Gerhard[4], Pruschy Martin[1], Guckenberger Matthias[1], Klöck Stephan[1], Riesterer Oliver[1]

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[2] Department of Nuclear Medicine, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[3] Department of Diagnostic and Interventional Radiology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[4] Department of Otorhynolaryngeology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[5] Institute of Clinical Pathology, University Hospital Zurich and University of Zurich, Zurich Switzerland.

Status of the manuscript:

published in Radiother Oncol. 2015;117(1):125-131.

Author contribution Marta Bogowicz (Marta Nesteruk):

- Development of a DICOM-compatible software implementation for perfusion calculation
- Full data analysis (results for Tables 2-4 and Figures 1-2) and interpretation
- Preparation of all supplementary files
- Manuscript writing, revision, editing and proofreading

Radiotherapy and Oncology 117 (2015) 125–131

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

FDG PET in head and neck

Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma

Marta Nesteruk [a,*], Stephanie Lang [a], Patrick Veit-Haibach [b,c], Gabriela Studer [a], Sonja Stieb [a], Stefan Glatz [a], Hossein Hemmatazad [a], Kristian Ikenberg [e], Gerhard Huber [d], Martin Pruschy [a], Matthias Guckenberger [a], Stephan Klöck [a], Oliver Riesterer [a]

[a] *Department of Radiation Oncology;* [b] *Department of Nuclear Medicine;* [c] *Department of Diagnostic and Interventional Radiology;* [d] *Department of Otorhynolaryngeology; and* [e] *Institute of Clinical Pathology, University Hospital Zurich, University of Zurich, Switzerland*

ARTICLE INFO

Article history:
Received 18 July 2015
Received in revised form 24 August 2015
Accepted 17 September 2015

Keywords:
Head and neck cancer
FDG PET
CT perfusion

ABSTRACT

Background and purpose: This study investigated whether tumor perfusion, FDG uptake and their correlation depend on tumor stage, site and HPV in head and neck cancer.
Material and methods: 41/55 eligible patients with integrated FDG-PET/perfusion CT from 2 prospective studies were assessed. A GTV^{CT} and GTV^{PET} were created. Perfusion maps were calculated using singular value decomposition method. Blood volume (BV), blood flow (BF), mean transit time (MTT) and standardized uptake value (SUV) in the tumor were compared to the surrounding tissue using Wilcoxon test and Spearman correlation of perfusion and SUV_{mean} in the tumor was studied ($p = 0.05$).
Results: Perfusion parameters were significantly increased in the GTV^{CT} of advanced tumors in comparison to the surrounding soft tissue ($p < 0.01$). Oral cavity and oropharyngeal cancer showed a higher BF than laryngeal cancer ($p < 0.04$). No correlation between perfusion and SUV_{mean} was found, however SUV_{mean} correlated significantly with BF for the HPV-positive tumors ($r = 0.86$, $p = 0.04$) and with BV for the oropharyngeal cancer ($r = 0.63$, $p = 0.05$).
Conclusion: Tumor stage, site and HPV are associated with different perfusion or combined perfusion/SUV signatures. Further studies are needed to investigate if these signatures co-determine clinical outcome.

The response of human head and neck squamous cell carcinoma (HNSCC) to radiotherapy depends not only on tumor size but also on tumor site and biology [1,2]. Results from clinical trials indicate that patients with oropharyngeal carcinoma (OPC) have the highest loco-regional control rate followed by laryngeal, hypopharyngeal and oral cavity cancers [3,4]. Part of this differential outcome is explained by the high rate of human papilloma virus (HPV)-induced OPC. Since the knowledge about HPV as prognostic and predictive marker is rather new we are only beginning to understand the biology of HPV associated cancers. Recent studies using in vivo tumor models suggest that HPV increases radiosensitivity [7] and may also interfere with the tumor microenvironment, e.g. HPV was associated with a more rapidly decreasing hypoxic fraction in HPV-positive HNSCC xenografts in comparison to HPV-negative tumors [8] and xenografts formed by HPV-positive cells had fewer hypoxic areas and displayed a lower hypoxic signature [9].

In addition to HPV infection tumor size and tumor site specific anatomy might co-determine tumor physiology as well. For example large tumors might be more heterogenous with regard to vascularization than smaller tumors due to necrotic areas. In addition, cancers of the tongue and floor of mouth grow in an anatomical region with abundant normal vasculature and therefore might display increased angiogenesis [10] in comparison to cancers of the larynx.

A surrogate marker for tumor angiogenesis is tumor perfusion assessed by computed tomography perfusion (CTP) [11–13]. Perfusion is mainly described by three parameters: blood flow (BF), blood volume (BV) and mean transit time (MTT) [14]. BF is defined as the rate of blood supplying a particular area. BV represents the volume of blood within the vasculature. The average time taken by blood to travel from the arterial to the venous end is called MTT. All three parameters might be differentially expressed due to a tumor

---

* Corresponding author at: University Hospital Zurich, Department of Radiation Oncology, Rämistrasse 100, CH-8091 Zurich, Switzerland.

E-mail address: Marta.Nesteruk@usz.ch (M. Nesteruk).

http://dx.doi.org/10.1016/j.radonc.2015.09.026

biology and site-specific vascularity and functionality of the vasculature.

In the clinic, the most important functional imaging parameter used in the diagnostic workup of HNSCC patients so far is 18F-fluorodeoxyglucose ([18F]-FDG) PET. [18F]-FDG uptake in the tumor is a balanced signal of tumor cell metabolism, hypoxia, perfusion and aggressiveness [15,16]. On the other hand, tumor vasculature is known to be disorganized, tortuous and leaky and thereby largely ineffective [17]. Here we hypothesized that tumor perfusion, [18F]-FDG uptake and their correlation differ dependent on tumor stage, site and HPV status.

Materials and methods

Study population

55 patients, who had a pre-treatment [18F]-FDG-PET with integrated perfusion CT in 2 separate prospective studies between 2008 and 2014, were pooled and assessed. Both studies had ethical approval. The patients had suspicion for head and neck cancer and were scanned before histological verification by panendoscopy. 14 patients had to be excluded from our analysis due to only inflammation present in the PET/CT scan ($n = 7$), non-squamous cell carcinoma ($n = 3$), no primary tumor inside the limited region of the CTP ($n = 2$), spine deformation ($n = 1$) or stenosis of carotid arteries ($n = 1$). Finally, 41 patients with HNSCC were enrolled (Table 1). HPV status was determined by p16 analysis ($n = 36$).

[18F]-FDG PET/CT and CT perfusion imaging

All images were acquired with a combined PET/CT scanner (Discovery VCT, GE Healthcare, Milwaukee, WI, USA). The imaging procedure is described in the Supplement. The region for the CTP examination was defined based on the [18F]-FDG uptake. The obtained images had the following characteristics: PET in-plane pixel size of 5.08 mm and slice thickness of 3.27 mm; CTP temporal resolution of 1 s (45 scans), in-plane pixel size of 0.7 mm, slice thickness of 5 mm and cranial-caudal coverage of 4 cm. As the PET and CTP images were taken consecutively no additional registration was performed.

Volumes contouring

The GTV^{CT} of the primary tumor was contoured on the perfusion CT, and all contours were checked by an experienced senior radiation oncologist. Auto-segmentation of the PET signal in the tumor (GTV^{PET}) was performed. A gradient-based segmentation method was chosen, as it better estimates a true tumor volume than a threshold approach in the case of laryngeal cancer [19] and performs equally good for other head and neck tumors [20]. PET images were pre-processed and segmented as described by Geets et al. [19] using the Python programming language (version 2.7.6).

Table 1
Patients included into the study.

Group		Number of patients
All patients		41
Tumor stage	T1/T2	12
	T3/T4	29
Tumor site	Oropharynx	17
	Hypopharynx	8
	Larynx	8
	Oral cavity	8
HPV status	Positive	9
	Negative	27
N stage	N0	12
	N1/N2/N3	29
Gender	Female	7
	Male	34
Age	Mean 60 years (43–79 years)	41
GTV volume	Mean 20.2 cm^3 (0.4–91.9 cm^3)	41
Study	2008–2010	27
	2012–2014	14

Quantitative image comparison and statistical methods

An in-house software for perfusion maps determination and image comparisons was developed. Perfusion calculations were based on the singular value decomposition method [21]. The final results were normalized to the perfusion in the carotid artery. The details of the perfusion parameter calculation are described in the Supplement. A window of −20 to 180 HU was used to exclude the bone, air and metal artifacts from the analysis. CTP image resolution was adjusted to the resolution of the PET image by signal averaging. The analysis was comprised of the following parts: The GTV^{CT} and the GTV^{PET} were compared using Dice's coefficient [22] and Hausdorff distance [23]. The mean perfusion parameters (blood volume BV, blood flow BF and mean transit time MTT) were determined inside the tumor and its surrounding tissue and their differences were examined using paired-samples Wilcoxon Signed-Rank test: the PET non-active part of the GTV^{CT} was compared to the active GTV^{PET} and the whole GTV^{CT} was compared to its surrounding normal tissue, which was a 10 mm ring structure around the tumor (ring structure M). Mean perfusion parameters and [18F]-FDG uptake values inside the GTV^{CT} were compared between tumor stage, site and HPV status using Kruskal–Wallis and Mann–Whitney tests. For each patient, standard deviations of the standardized uptake values (SUV) and the perfusion parameters in the different $GTVs^{CT}$ were calculated and normalized (nSD) to the mean value in this region to analyze the homogeneity inside the $GTVs^{CT}$. The SUV_{mean} and mean CTP signals inside the GTV^{PET} and the whole GTV^{CT} were correlated by Spearman correlation.

The analysis was performed for the whole patient population and subsequently for subgroups based on stratification into tumor stage, site and HPV status. For each of the subgroups the multiple testing correction was performed using Holm–Bonferroni method [24] with family-wise error rate of 0.05. Corrected p-values are denoted in this paper as p_c. The uncorrected strikingly low p-values were also mentioned ($p < 0.05$) as a hint for the future studies [25]. The statistical analysis was performed using Python library SciPy.

Results

GTV^{CT} and GTV^{PET} volumes' comparison

The mean GTV^{CT} and GTV^{PET} were 20.5 cm^3 (range 0.4–91.9 cm^3) and 14.2 cm^3 (0.5–67.9 cm^3), respectively. In one case the segmentation was not possible due to the low PET signal in the tumor (oropharynx T1 HPV-negative). The CT scan before the PET acquisition and perfusion CT were rigidly registered in the commercial software Eclipse (v. 11, Varian Medical Systems, Palo Alto, California, USA) to evaluate potential patient movement. The mean motion was 4.9 mm, which is smaller than the image resolution. One patient (hypopharynx T3 HPV negative) was excluded due to large motion (16.1 mm). An example of perfusion and SUV maps together with contours is presented in Fig. 1. The mean Dice's coefficient for GTV^{CT} and GTV^{PET} was 0.63 (range 0.23–0.84), whereas the mean Hausdorff distance was 23 mm

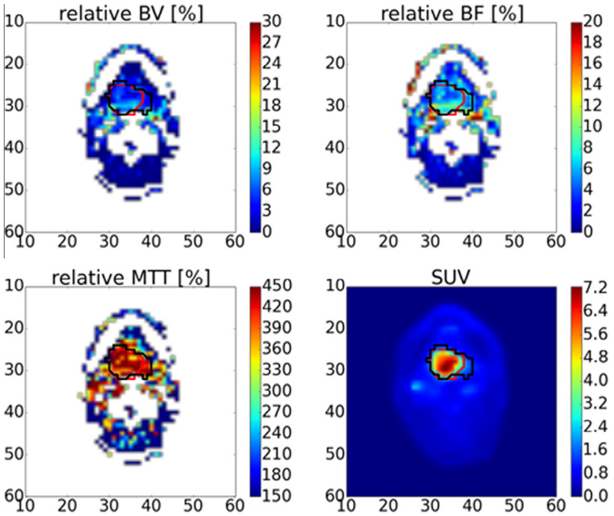


Fig. 1. Example of perfusion and standardized uptake value (SUV) maps for one of the patients. The red line corresponds to GTV^{PET} and black one to GTV^{CT}. The white pixels in the perfusion maps refer to the pixels with HU values outside the defined range. BV – blood volume, BF – blood flow, MTT – mean transit time.

(range 10–60 mm) indicating substantial heterogeneity and variability between the two volumes.

Perfusion parameters in the GTV^{CT} and GTV^{PET} in comparison to the surrounding normal tissue

BV and MTT perfusion parameters were significantly increased in the GTV^{CT} compared to the 1 cm ring structure M (BV difference = 26%, $p_c < 0.01$; MTT difference = 16%, $p_c < 0.01$). An increased tumor perfusion in comparison to the surrounding soft tissue was found for stages T3/T4 (BV difference = 31%, $p_c < 0.01$; MTT difference = 21%, $p_c < 0.01$; BF difference = 19%, $p_c = 0.01$) but not for T1/T2. Regarding HPV status increased perfusion in the GTV^{CT} was observed for the HPV negative (BV difference = 21%, $p_c < 0.006$: MTT difference = 16%, $p_c = 0.02$), but not for positive tumors. With regard to the [18F]-FDG-active (GTV^{PET}) and non-active parts of the GTV^{CT}, no difference of the perfusion parameters was found (Table 2).

Perfusion parameters and standardized uptake value dependent on tumor site, stage and HPV

Kruskal–Wallis test showed strikingly low *p*-values without multiple testing for the BF in the GTV^{PET} in different tumor sites ($p = 0.02$). Larynx tumors were characterized with lower BF in comparison to OPC ($p = 0.04$) and oral cavity ($p < 0.01$) tumors (Fig. 2). The higher BF was observed in the GTV^{PET} in HPV negative tumors ($p = 0.03$). These results were not significant after multiple corrections. All other parameters including SUV_{mean} did not differ according to tumor site and HPV status. With regard to tumor stage SUV_{mean} was higher for stages T3/T4 (GTV^{PET} $p_c = 0.02$; GTV^{CT} $p_c = 0.01$), while no difference was observed in the perfusion.

Inhomogeneity of CTP and PET parameters

The mean values of the nSDs were calculated to represent the inhomogeneity for the whole population. The largest inhomogeneity was found for BV (nSD = 71%) and the smallest for MTT (nSD = 34%). For BF and SUV nSD of 53% was found.

Correlation of mean perfusion parameters with SUV for the whole population

The correlations of MTT and SUV_{mean} had strikingly low p-values (GTV^{PET} $r = 0.40$, $p = 0.02$ and GTV^{CT} $r = 0.35$, $p = 0.04$). However after correcting for multiple testing, these correlations were not significant anymore. No other statistically significant correlation was found (Table 3).

Correlation of mean perfusion parameters with SUV dependent on tumor stage, tumor site and HPV

Significant correlation was found after multiple testing regarding the HPV status and tumor site (Table 3). The SUV_{mean} correlated in the GTV^{PET} with BF for HPV positive tumors ($r = 0.86$, $p_c = 0.04$) and in the GTV^{CT} with BV for OPC ($r = 0.63$, $p_c = 0.05$). In addition, several in part differential correlations with strikingly low *p*-values without multiple testing were observed. The SUV_{mean} correlated with BV for the larynx (GTV^{PET} $r = 0.83$, $p = 0.04$), stage T3/T4 (GTV^{PET} $r = 0.42$, $p = 0.05$) and HPV negative (GTV^{PET} $r = 0.44$; $p = 0.05$; GTV^{CT} $r = 0.51$, $p = 0.01$) tumors. The HPV negative tumors showed also a correlation of SUV_{mean} with MTT (GTV^{PET} $r = 0.46$, $p = 0.04$). The negative correlation of SUV_{mean} and BF was found for the laryngeal cancer (GTV^{CT} $r = -0.75$, 0.05).

Discussion

Using an integrated PET/CT scanner this study analyzed tumor stage, tumor site and HPV dependent tumor perfusion and [18F]-FDG activity as well as the correlation of these parameters. Our results show that perfusion in the tumor is significantly different in comparison to the surrounding soft tissue. This difference was highly significant for advanced tumors (T3/T4) but not for small tumors (T1/T2). With regard to tumor site, only BF differed with laryngeal cancer having a lower BF compared to OPC and oral cavity cancer. Mean BF was also decreased in the HPV-negative tumors in comparison to the positive ones. In addition, we did not find significant correlation of perfusion parameters and [18F]-FDG uptake for the whole population, however we observed significant correlation dependent on tumor site and HPV status.

The prognosis of HNSCC differs according to tumor site. In a recent phase III randomized trial, which compared different fractionation regimens, patients with OPC had a LCR of 88% after 5 years in comparison to only 49% for oral cavity carcinoma [3]. In this trial, patients with HPV-negative OPC had a LCR of 74%, which was still significantly better in comparison to oral cavity cancers, which are HPV-negative in 90–95% of cases. The reason why HPV-positive cancers respond better is not yet completely understood and has been linked to differential intrinsic radiosensitivity [7] as well as to altered amount and behavior of tumor hypoxia during radiotherapy in experimental models [9,8]. In addition to HPV status, factors related to the anatomical site co-determine treatment outcome, e.g. tumor physiological factors such as the possibility to recruit blood vessels for regional or distant spread of tumor cells or, on the contrary, a lack of regional blood supply resulting in tumor hypoxia. This is supported, for example, by anatomical studies showing a high vascular supply to cancers of the tongue [10].

It is well known that vasculature differs between tumor and healthy tissue, e.g. inside solid tumors angiogenesis is increased, but on the other hand tumor vasculature is disorganized, tortuous and leaky [17]. Increased tumor perfusion indicates increased angiogenesis. This is supported by clinical studies in patients with prostate and colorectal cancer as well as preclinical studies using VX2 lung tumors implanted in mice showing a correlation between perfusion parameters and microvessel density [11–13]. The

Table 2
Median differences of mean perfusion parameters for the FDG-active ($\mathrm{GTV}^{\mathrm{PET}}$) and non-active parts of the tumor ($\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$) as well as the $\mathrm{GTV}^{\mathrm{CT}}$ and its margin. BV – blood volume, BF – blood flow, MTT – mean transit time, M – 10 mm ring structure around $\mathrm{GTV}^{\mathrm{CT}}$, d – difference of relative perfusion parameters normalized to the $\mathrm{GTV}^{\mathrm{CT}}$, p – significance level, p_c – significance level corrected for the multiple testing. Differences with strikingly low p-values ($p < 0.05$) are underlined and statistically significant values ($p_c < 0.05$) are in bold.

Tumor subgroup	Comparison		Perfusion parameters		
	Structure 1	Structure 2	BV	MTT	BF
All $n = 39$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 7.0\%$ $p = 0.62$ $p_c = 1.25$	$d = 3.9\%$ $p = 0.393$ $p_c = 1.18$	$d = 3.4\%$ $p = 0.67$ $p_c = 0.67$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 26\%$ $\underline{p < 0.01}$ $\mathbf{p_c < 0.01}$	$d = 16\%$ $\underline{p < 0.01}$ $\mathbf{p_c < 0.01}$	$d = 16\%$ $\underline{p = 0.02}$ $p_c = 0.07$
T1/T2 $n = 11$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 2.0\%$ $p = 0.96$ $p_c = 1.92$	$d = 8.5\%$ $p = 0.40$ $p_c = 1.99$	$d = -3.8\%$ $p = 0.61$ $p_c = 1.84$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 18\%$ $\underline{p = 0.05}$ $p_c = 0.30$	$d = 3.8\%$ $p = 0.58$ $p_c = 2.30$	$d = -0.80\%$ $p = 1.00$ $p_c = 1.00$
T3/T4 $n = 28$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 9.4\%$ $p = 0.49$ $p_c = 0.97$	$d = 3.9\%$ $p = 0.61$ $p_c = 0.61$	$d = 7.6\%$ $p = 0.48$ $p_c = 1.43$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 31\%$ $\underline{p < 0.01}$ $\mathbf{p_c < 0.01}$	$d = 21\%$ $\underline{p < 0.01}$ $\mathbf{p_c < 0.01}$	$d = 19\%$ $\underline{p < 0.01}$ $\mathbf{p_c = 0.01}$
Oropharynx $n = 16$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = -1.3\%$ $p = 0.78$ $p_c = 0.78$	$d = 3.6\%$ $p = 0.45$ $p_c = 0.89$	$d = -5.8\%$ $p = 0.29$ $p_c = 1.43$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 23\%$ $\underline{p = 0.02}$ $p_c = 0.09$	$d = 5.5\%$ $p = 0.31$ $p_c = 1.26$	$d = 3.6\%$ $p = 0.44$ $p_c = 1.32$
Hypopharynx $n = 7$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 21\%$ $p = 0.50$ $p_c = 1.50$	$d = 4.6\%$ $p = 0.50$ $p_c = 0.50$	$d = 17\%$ $p = 0.50$ $p_c = 1.00$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 28\%$ $p = 0.12$ $p_c = 0.46$	$d = 21\%$ $p = 0.08$ $p_c = 0.40$	$d = 14\%$ $p = 0.08$ $p_c = 0.48$
Larynx $n = 8$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = -7.0\%$ $p = 0.35$ $p_c = 1.04$	$d = -8.9\%$ $p = 0.35$ $p_c = 0.69$	$d = 5.0\%$ $p = 0.46$ $p_c = 0.46$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 31\%$ $\underline{p = 0.02}$ $p_c = 0.11$	$d = 28\%$ $\underline{p = 0.04}$ $p_c = 0.17$	$d = 19\%$ $\underline{p = 0.04}$ $p_c = 0.22$
Oral cavity $n = 8$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 24\%$ $p = 0.24$ $p_c = 0.47$	$d = 5.7\%$ $p = 0.13$ $p_c = 0.38$	$d = 11\%$ $p = 0.13$ $p_c = 0.51$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 29\%$ $\underline{p = 0.03}$ $p_c = 0.17$	$d = 14\%$ $\underline{p = 0.05}$ $p_c = 0.23$	$d = -1.9\%$ $p = 0.92$ $p_c = 0.92$
HPV-positive $n = 9$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = -8.2\%$ $p = 0.40$ $p_c = 1.20$	$d = 1.3\%$ $p = 0.89$ $p_c = 0.89$	$d = -5.8\%$ $p = 0.26$ $p_c = 1.05$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 29\%$ $\underline{p = 0.02}$ $p_c = 0.10$	$d = -0.72\%$ $p = 0.69$ $p_c = 1.37$	$d = 21\%$ $\underline{p = 0.04}$ $p_c = 0.22$
HPV-negative $n = 26$	$\mathrm{GTV}^{\mathrm{PET}}$	$\mathrm{GTV}^{\mathrm{CT}} - \mathrm{GTV}^{\mathrm{PET}}$	$d = 7.0\%$ $p = 0.60$ $p_c = 0.60$	$d = 4.4\%$ $p = 0.26$ $p_c = 1.02$	$d = 9.2\%$ $p = 0.47$ $p_c = 0.94$
	$\mathrm{GTV}^{\mathrm{CT}}$	M	$d = 21\%$ $\underline{p < 0.01}$ $\mathbf{p_c < 0.01}$	$d = 16\%$ $\underline{p < 0.01}$ $\mathbf{p_c = 0.02}$	$d = 2.4\%$ $p = 0.31$ $p_c = 0.92$

increased BV in the tumor shown by us is consistent with other studies [18,26]. However, whereas all other studies compared perfusion in the tumor with either muscle or contra-lateral healthy tissue, we took the immediate surrounding soft tissue for comparison. The rationale for using the surrounding soft tissue was the limited scan region of 4 cm, which precludes comparison of tumors in different sites with the same reference muscle. In addition the comparison of tumor tissue with the normal soft tissue from where the tumor originated eliminates the influence of differential vascularization in different anatomical regions. A counter argument for our approach would be that angiogenesis might also be increased in the immediate surroundings of the tumor due to microscopic tumor extension. Mean pathologic extension in HNSCC was shown to be below 5 mm in a pathological study [27]. The use of different normal tissue regions for comparison of perfusion values is one possible explanation why, in contrast to

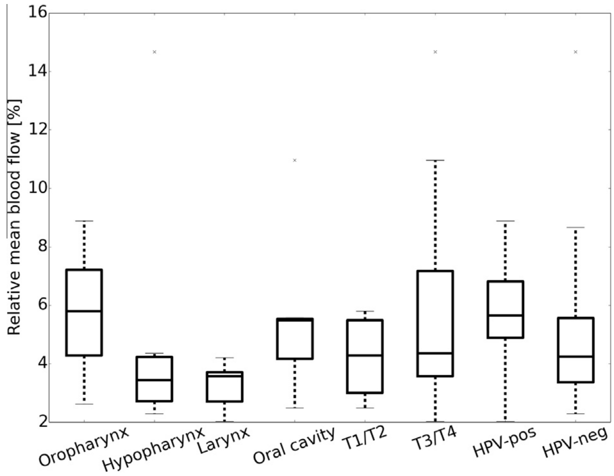


Fig. 2. Boxplots showing the mean values of the blood flow inside the GTV^{PET} for different tumor sites and tumor stages.

our observation of increased MTT, other investigators found decreased MTT in the tumor [11,18].

Interestingly, we did not find significant differences in mean perfusion parameters of large versus small tumors. It would be interesting to perform heterogeneity analyses in future studies based on the assumption of higher probability of tumor necrosis and hypoxia in larger tumors.

We showed significant correlation of tumor metabolism and perfusion. This is expected from the existing literature investigating the underlying biology of tumor metabolism and [18F]-FDG uptake [15,16]. Several studies in head and neck, lung, breast and rectal cancers investigated correlation of [18F]-FDG and CT perfusion parameters [28–33]. These studies used different perfusion models and parameters. So far only few studies investigated the correlation between CT perfusion and [18F]-FDG metabolism in head and neck cancer: a summary of the correlations between perfusion and SUV parameters found by us and in other studies of HNSCC is presented in Table 4. The overview is limited to the perfusion parameters studied by us. Patient numbers in these studies were small, no clinically relevant subgroup analyses were made and the results presented are inconclusive. Our subgroup analysis in terms of tumor HPV status provides an additional explanation to contradictory results of previously published studies regarding SUV correlations with BF [18,26,34]. We did not observe a correlation for the whole patient cohort, but we found a significant correlation for the HPV positive tumors.

Strikingly low *p*-values, however not significant after multiple testing corrections, were shown for the correlations of the BV and SUV_{mean} in the case of HPV negative tumors as well as T3/T4 stages. Also the differences in the BV and MTT between the GTV^{CT} and its 1 cm margin were statistically significant for the T3/T4 and HPV-negative subgroups. As these groups are associated with impaired prognosis it might be a signature for a high risk tumor physiology, e.g. pro-angiogenic, ineffective vasculature and pro-metastatic tumor bed conditions. In this study the size of these subgroups was larger in comparison to T1/T2 and HPV-positive subgroups, which could also influence the obtained results. Further studies on larger, equally stratified patient cohorts are needed to validate these correlations as well as studies in form of correlations

Table 3

Spearman correlation of perfusion parameters and SUV_{mean} according to tumor stage, tumor site and HPV status. BV – blood volume, BF – blood flow, MTT – mean transit time, GTV^{PET} – GTV auto-segmented based on the PET signal, *r* – correlation coefficient, *p* – significance level, p_c – significance level corrected for the multiple testing. Correlations with strikingly low *p*-values ($p < 0.05$) are underlined and statistically significant values ($p_c < 0.05$) are in bold.

Tumor subgroup	Correlation of Perfusion Parameters and SUV_{mean}					
	BV		BF		MTT	
	GTV^{PET}	GTV^{CT}	GTV^{PET}	GTV^{CT}	GTV^{PET}	GTV^{CT}
All	$r = 0.29$ p = 0.09 $p_c = 0.17$	$r = 0.29$ $p = 0.08$ $p_c = 0.25$	$r = 0.34$ $\underline{p = 0.05}$ $p_c = 0.21$	$r = 0.21$ $p = 0.22$ $p_c = 0.22$	$r = 0.40$ $\underline{p = 0.02}$ $p_c = 0.12$	$r = 0.35$ $\underline{p = 0.04}$ $p_c = 0.19$
T1/T2	$r = -0.02$ $p = 0.96$ $p_c = 0.96$	$r = 0.53$ $p = 0.08$ $p_c = 0.45$	$r = 0.38$ $p = 0.31$ $p_c = 0.96$	$r = 0.49$ $p = 0.13$ $p_c = 0.63$	$r = 0.27$ $p = 0.49$ $p_c = 0.98$	$r = 0.47$ $p = 0.14$ $p_c = 0.57$
T3/T4	$r = 0.42$ $\underline{p = 0.05}$ $p_c = 0.29$	$r = 0.20$ $p = 0.34$ $p_c = 0.69$	$r = 0.30$ $p = 0.16$ $p_c = 0.64$	$r = 0.17$ $p = 0.45$ $p_c = 0.45$	$r = 0.38$ $p = 0.07$ $p_c = 0.37$	$r = 0.28$ $p = 0.20$ $p_c = 0.60$
Oropharynx	$r = 0.31$ $p = 0.26$ $p_c = 0.52$	$r = 0.63$ $\underline{p = 0.01}$ $\mathbf{p_c = 0.05}$	$r = 0.45$ $p = 0.12$ $p_c = 0.49$	$r = 0.48$ $p = 0.08$ $p_c = 0.42$	$r = 0.32$ $p = 0.29$ $p_c = 0.29$	$r = 0.37$ $p = 0.20$ $p_c = 0.590$
Hypopharynx	$r = 0.60$ $p = 0.21$ $p_c = 1.04$	$r = 0.66$ $p = 0.16$ $p_c = 0.94$	$r = 0.26$ $p = 0.62$ $p_c = 1.25$	$r = 0.37$ $p = 0.47$ $p_c = 1.41$	$r = 0.43$ $p = 0.40$ $p_c = 1.59$	$r = -0.03$ $p = 0.96$ $p_c = 0.96$
Larynx	$r = 0.83$ $\underline{p = 0.04}$ $p_c = 0.25$	$r = -0.36$ $p = 0.43$ $p_c = 0.86$	$r = -0.54$ $p = 0.27$ $p_c = 0.80$	$r = -0.75$ $\underline{p = 0.05}$ $p_c = 0.26$	$r = 0.71$ $p = 0.11$ $p_c = 0.44$	$r = 0.32$ $p = 0.48$ $p_c = 0.48$
Oral cavity	$r = -0.36$ $p = 0.43$ $p_c = 1.73$	$r = -0.21$ $p = 0.64$ $p_c = 1.93$	$r = -0.07$ $p = 0.88$ $p_c = 1.76$	$r = 0.071$ $p = 0.88$ $p_c = 0.88$	$r = 0.46$ $p = 0.29$ $p_c = 1.47$	$r = 0.50$ $p = 0.25$ $p_c = 1.52$
HPV-positive	$r = 0.48$ $p = 0.23$ $p_c = 1.17$	$r = -0.19$ $p = 0.65$ $p_c = 1.95$	$r = 0.86$ $\underline{p = 0.01}$ $\mathbf{p_c = 0.04}$	$r = 0.33$ $p = 0.42$ $p_c = 1.68$	$r = 0.10$ $p = 0.82$ $p_c = 1.65$	$r = -0.02$ $p = 0.96$ $p_c = 0.96$
HPV-negative	$r = 0.44$ $\underline{p = 0.05}$ $p_c = 0.18$	$r = 0.51$ $\underline{p = 0.01}$ $p_c = 0.08$	$r = 0.33$ $p = 0.16$ $p_c = 0.31$	$r = 0.19$ $p = 0.40$ $p_c = 0.40$	$r = 0.46$ $\underline{p = 0.04}$ $p_c = 0.20$	$r = 0.41$ $p = 0.06$ $p_c = 0.19$

Table 4
Summary of published perfusion and SUV correlations in HNSCC. Only parameters studied in this paper are shown.

Study	Perfusion parameter	SUV	Correlation coefficient	Primary tumors	Perfusion calculation method
Hirasawa et al. [34]	BF	Mean	−0.538	Head and neck (squamous cell carcinoma, poorly differentiated carcinoma, adenocarcinoma, follicular lymphoma	Maximum slope
Bisdas et al. [26]	BF	Mean Max	0.57 0.63	Head and neck (oropharynx, larynx, skin, hypopharynx, nasopharynx)	Singular value decomposition
Veit-Haibach et al. [18]	–	–	No correlation for primary tumor	Head and neck (oropharynx, hypopharynx, larynx, oral cavity, sinonasal, a squamous cell cancer of the mandible)	Singular value decomposition
This study	BF BV	Mean Mean	0.86 0.63	HPV-positive oropharynx	Singular value decomposition

of tissue and imaging markers to provide a better understanding of the perfusion parameters on the biological level.

In contrast to tumor biopsies that only provide a snapshot of a small tumor area, functional imaging allows delineating tumor biology and physiology with spatial resolution. This is the prerequisite for the development of dose painting, i.e. the distribution of radiation dose within the tumor dependent on radiation sensitivity. Due to the broad availability of integrated PET/CT scanners the combination of [18F]-FDG PET and functional CT is promising for future clinical research in this direction.

There are also some limitations of the imaging procedures used in our study. An important issue is patient fixation during imaging. The large inhomogeneity of perfusion parameters and SUV in the $\mathrm{GTV}^{\mathrm{CT}}$ suggest that future studies should best be performed by voxel-wise correlations. However, this requires very stable patient positioning with possible movement smaller than image resolution, which only can be achieved by use of a thermoplastic mask during the 30 min lasting PET/CTP procedure.

The auto-segmented $\mathrm{GTV}^{\mathrm{PET}}$ was on average 65% overlapping with the $\mathrm{GTV}^{\mathrm{CT}}$ contoured by the radiation oncologists, which is in agreement with the published data [35]. This mismatch is explained by the known fact that often only part of the tumor is [18F]-FDG-active and by the motion of the patient.

In summary, we found different perfusion in the tumor in comparison to the surrounding soft tissue as well as differential correlation of perfusion and PET parameters dependent on HPV status. Further studies on tumor physiology in HNSCC should therefore take into account the various subgroups. A significantly different perfusion or combined perfusion/SUV signature dependent on the subgroup might predict outcome in HNSCC. Further studies are needed with outcome analyses of large cohorts as well as studies that correlate perfusion parameters with immunohistochemical markers. In addition, the inhomogeneity of tumors in terms of vascularity and metabolism requires voxel-wise correlation studies with proper patient fixation during the imaging procedure. In the long-term this research points to a better understanding of differential tumor response of HNSCC.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Acknowledgments

The project was supported by the KFSP Tumor Oxygenation of the University of Zurich, by a grant of the Center for Clinical Research, University and University Hospital Zurich and by a research grant from Merck (Schweiz) AG.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2015.09.026.

References

[1] Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. Science 2001;333:1157–60.

[2] Chung CH, Zhang Q, Kong CS, et al. P16 protein expression and human papillomavirus status as prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. J Clin Oncol 2014;32:3930–8.

[3] Bentzen J, Toustrup K, Eriksen JG, Primdahl H, Andersen LJ, Overgaard J. Locally advanced head and neck cancer treated with accelerated radiotherapy, the hypoxic modifier nimorazole and weekly cisplatin. Results from the DAHANCA 18 phase II study. Acta Oncol 2015;54:1–7.

[4] Studer G, Luetolf UM, Glanzmann C. Locoregional failure analysis in head-and-neck cancer patients treated with IMRT. Strahlenther Onkol 2007;1:417–23.

[7] Sørensen BS, Busk M, Olthof N, Speel EJ, Horsman MR, Alsner J, et al. Radiosensitivity and effect of hypoxia in HPV positive head and neck cancer cells. Radiother Oncol 2013;1:500–5.

[8] Sørensen BS, Busk M, Horsman MR, Alsner J, Overgaard J, Kyle AH, et al. Effect of radiation on cell proliferation and tumor hypoxia in HPV-positive head and neck cancer in vivo models. Anticancer Res 2014;34:6297–304.

[9] Hanns E, Job S, Coliat P, et al. Human Papillomavirus-related tumours of the oropharynx display a lower tumour hypoxia signature. Oral Oncol 2015;51:848–56.

[10] Wistar C. A System of Anatomy: For the Use of Students of Medicine. 9th ed. Philadelphia: Thomas, Cowperthwait & Co.; 1846.

[11] Luczynska E, Gasinska A, Blecharz P, Stelmach A, Jereczek-Fossa BA, Reinfuss M. Value of perfusion CT parameters, microvessel density and VEGF expression in differentiation of benign and malignant prostate tumours. Pol J Pathol 2014;65:229–36.

[12] Goh V, Halligan S, Daley F, Wellsted DM, Guenther T, Bartram CI. Colorectal tumor vascularity: quantitative assessment with multidetector CT– do tumor perfusion measurements reflect angiogenesis? Radiology 2008;249:510–7.

[13] Ling S, Deng D, Mo Y, Zhang X, Guan X, Wei Q. Correlations between CT perfusion parameters and vascular endothelial growth factor expression and microvessel density in implanted VX2 lung tumors. Cell Biochem Biophys 2014;70:629–33.

[14] Lee TY, Purdie TG, Stewart E. CT imaging of angiogenesis. Q J Nucl Med 2010;41:171–87.

[15] Rajendran JG, Mankoff DA, O'Sullivan F, et al. Hypoxia and glucose metabolism in malignant tumors: evaluation by [18F]fluoromisonidazole and [18F] fluorodeoxyglucose positron emission tomography imaging. Clin Cancer Res 2004;10:2245–52.

[16] Warburg O. On the origin of cancer cells. Science 1956;1:309–14.

[17] Chung AS, Lee J, Ferrara N. Targeting the tumour vasculature: insights from physiological angiogenesis. Nat Rev Cancer 2010;10:505–14.

[18] Veit-Haibach P, Schmid D, Strobel K, et al. Combined PET/CT-perfusion in patients with head and neck cancers. Eur Radiol 2013;23:163–73.

[19] Geets X, Lee JA, Bol A, Lonneux M, Grégoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging 2007;34:1427–38.

[20] Chen H, Jiang J, Gao J, Liu D, Axelsson J, Cui M, et al. Tumor volumes measured from static and dynamic 18F-fluoro-2-deoxy-D-glucose positron emission tomography-computed tomography scan: comparison of different methods using magnetic resonance imaging as the criterion standard. J Comput Assist Tomogr 2014;38:209–15.

[21] Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part I: mathematical approach and statistical analysis. Magn Reson Med 1996;36:715–25.
[22] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:197–302.
[23] Rockafellar RT, Wets RJ-B. Variational Analysis. Berlin: Springer-Verlag; 2005.
[24] Holm S. A simple sequentially rejective multiple test procedure. Scand J Statistics 1979;6:65–70.
[25] Victor A, Elsäßer A, Hommel G, Blettner M. Judging a plethora of *p*-values. Dtsch Arztebl Int 2012;1:50–6.
[26] Bisdas S, Spicer K, Rumboldt Z. Whole-tumor perfusion CT parameters and glucose metabolism measurements in head and neck squamous cell carcinomas: a pilot study using combined positron-emission tomography/CT imaging. AJNR Am J Neuroradiol 2008;29:1376–81.
[27] Fleury B, Thariat J, Barnoud R, et al. Microscopic extensions of head and neck squamous cell carcinomas: impact for clinical target volume definition. Cancer Radiother 2014;18:666–71.
[28] van Elmpt W, Das M, Hüllner M, et al. Characterization of tumor heterogeneity using dynamic contrast enhanced CT and FDG-PET in non-small cell lung cancer. Radiother Oncol 2013;109:65–70.
[29] Sauter AW, Spira D, Schulze M, et al. Correlation between [18F]FDG PET/CT and volume perfusion CT in primary tumours and mediastinal lymph nodes of non-small-cell lung cancer. Eur J Nucl Med Mol Imaging 2013;40:677–84.
[30] Tixier F, Groves AM, Goh V. Correlation of intra-tumor 18F-FDG uptake heterogeneity indices with perfusion CT derived parameters in colorectal cancer. PLoS One 2014;9.
[31] Janssen MH, Aerts HJ, Buijsen J, Lambin P, Lammering G, Öllers MC. Repeated positron emission tomography-computed tomography and perfusion-computed tomography imaging in rectal cancer: fluorodeoxyglucose uptake corresponds with tumor perfusion. Int J Radiat Oncol Biol Phys 2012;82:849–55.
[32] Groves AM, Wishart GC, Shastry M, et al. Metabolic–flow relationships in primary breast cancer: feasibility of combined PET/dynamic contrast-enhanced CT. Eur J Nucl Med Mol Imaging 2009;36:416–21.
[33] Ippolito D, Capraro C, Guerra L, De Ponti E, Messa C, Sironi S. Feasibility of perfusion CT technique integrated into conventional 18FDG/PET-CT studies in lung cancer patients: clinical staging and functional information in a single study. Eur J Nucl Med Mol Imaging 2013;40:156–65.
[34] Hirasawa S, Tsushima Y, Takei H, et al. Inverse correlation between tumor perfusion and glucose uptake in human head and neck tumors. Acad Radiol 2007;14:312–8.
[35] Wanet M, Lee JA, Weynand B, De Bast M, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: A comparison with threshold-based approaches CT and surgical specimens. Radiother Oncol 2011;98:117–25.

Further reading

[5] Rischin D. Oropharyngeal cancer, human papilloma virus, and clinical trials. J Clin Oncol 2010;1:1–3.
[6] Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363:24–35.
[36] Yazdi M, Lari MA, Bernier G, Beaulieu L. An opposite view data replacement approach for reducing artifacts due to metallic dental objects. Med Phys 2011;38:2275–81.
[37] Abdoli M, Ay MR, Ahmadian A, Dierckx RA, Zaidi H. Reduction of dental filling metallic artifacts in CT-based attenuation correction of PET data using weighted virtual sinograms optimized by a genetic algorithm. Med Phys 2010;37:6166–77.

Supplement

[18F]-FDG PET/CT and CT perfusion imaging

After the measurement of blood sugar level, an activity of 270-410 MBq of [18F]-FDG was injected intravenously. The administered activity depended on the patient's weight. The imaging procedure started 60 minutes after [18F]-FDG injection with a CT examination for attenuation correction using the following scanning parameters: 80 mA, 140 kV, 0.5 s tube rotation and 4.25 mm slice thickness. In the next step the PET acquisition was performed for 2 minutes with a 15 cm axial field-of-view at each bed position. Total acquisition time of the PET was 12-16 min. Images were reconstructed with a 3D iterative algorithm with an in-plane pixel size and the slice thickness of the PET/CT images of 5.08 mm and 3.27 mm, respectively. The region for further CTP examination was defined based on the [18F]- FDG uptake.

The CTP scanning protocol suggested by the vendor was used. 40 ml of contrast media (Ultravist 370, Bayer Schering Healthcare, Germany) was injected at a flow of 5 ml/s in a cubital vein. After a delay of 5 s, the perfusion scans were acquired for 45 s (1 s rotation time with 1 image/s, 8 slices, 5 mm slice thickness, 80 mA, 100 kV) [1]. It enabled an anatomical cranial-caudal coverage of 4 cm of the target lesion. CT images were reconstructed using filtered back projection with an in-plane pixel size of the of 0.7 mm and the slice thickness of 5 mm.

Perfusion maps calculations

An in-house software for perfusion maps determination and image comparisons was developed in the Python programming language (version 2.7.6). To calculate perfusion maps the baseline image, arterial input function (AIF) and concentration of contrast agent in each voxel C_v were defined.

Based on the signal enhancement in an artery the time point of contrast arrival to the scanned region t_0 was defined individually for each patient. The baseline image was calculated as an average of images before t_0. Concentration of the contrast agent in each voxel as a function of the time $C_v(t)$ was calculated by subtracting the baseline image from the image at a certain time point.

For defining the AIF it is important to reduce partial volume effect. It was performed by choosing the artery with the biggest diameter - carotid artery. As the signal enhancement in the artery C_a did not differ throughout the scanned region, it was sufficient to contour a region of interest inside the artery in 2 to 3 slices. For the final AIF definition, the gamma function was fitted to the mean $C_a(t)$. The onset for recirculation was defined, by fitting the function only to the first pass of the bolus.

To speed up the calculations and reduce the noise, the resolution of perfusion maps was set to 2.8 mm by grouping the 4x4 pixels. A threshold of -20 to 180 Hounsfield units was defined, to exclude the bone, air, large vessels and metal artifacts from the analysis.

Blood volume was calculated as the total enhancement in the voxel C_v divided by the total enhancement in the artery C_a [2]:

$$BV = \frac{\int_0^\infty C_v(t)dt}{\int_0^\infty C_a(t)dt} \tag{1}$$

To obtain BF and MTT maps a deconvolution was performed using singular value decomposition (SVD) method [3]. The threshold of 20% of the maximum singular value in the matrix was introduced to reduce the influence of noise on the final results [4]. The adapted BF and MTT definitions were used, to account for the delayed arrival of the contrast at a certain voxel in comparison to the artery [4]:

$$BF = \max\big(k(t)\big)$$
$$MTT = \frac{1}{max\big(k(t)\big)} \int_0^\infty k(t)dt \tag{2}$$

where: k(t) is the flow scaled residual function.

All parameters were normalized to the mean values in the carotid artery.

References

[1] Veit-Haibach P, Schmid D, Strobel K et al. Combined PET/CT-perfusion in patients with head and neck cancers. Eur Radiol 2013;23:163-173.

[2] Konstas AA, Goldmakher GV, Lee TY, Lev MH. Theoretic Basis and Technical Implementations of CT Perfusion in Acute Ischemic Stroke, Part 1: Theoretic Basis. AJNR Am J Neuroradiol 2009;30:662-68.

[3] Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High Resolution Measurement of Cerebral Blood Flow using Intravascular Tracer Bolus Passages. Part I: Mathematical Approach and Statistical Analysis. Magn Reson Med 1996;36;715-25.

[4] Fieselmann A, Kowarschik M, Ganguly A, Hornegger J, Fahrig R. Deconvolution-Based CT and MR Brain Perfusion Measurement: Theoretical Model Revisited and Practical Implementation Details. International Journal of Biomedical Imaging 2011;2001.

3.2 Optimizing a Perfusion CT Protocol for Head and Neck Cancer

Stüssi Anja[1]*, Bogowicz Marta[1]*, Weichselbaumer Verena[2], Veit-Haibach Patrick[2], Riesterer Oliver[1], Guckenberger Matthias[1], Tanadini-Lang Stephanie[1]

* authors contributed equally

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[2] Department of Nuclear Medicine, University Hospital Zurich and University of Zurich, Zurich Switzerland.

Status of the manuscript:

published in Current Directions in Biomedical Engineering 2017, in press

Author contribution Marta Bogowicz:

- Phantom preparation and dose measurements
- Adaptation of scanning protocol
- Signal to noise analysis (results for Figure 3)
- Interpretation of the results
- Manuscript editing and proofreading

DE GRUYTER Current Directions in Biomedical Engineering 2017; 3(2): 591–594

Anja Stüssi*, Marta Bogowicz, Verena Weichselbaumer, Patrick Veit-Haibach, Oliver Riesterer, Matthias Guckenberger and Stephanie Tanadini-Lang

Optimizing a perfusion CT protocol for head and neck cancer

Abstract: Perfusion computed tomography (CTP) images tumor angiogenesis and can assess tumor aggressiveness. However, the CTP examinations are dose intensive. This study aimed to optimize a routinely used CTP protocol for the head and neck region in oncology in order to reduce the effective dose to the patient and simultaneously achieve the same image quality.

The Alderson phantom was scanned on a GE Revolution CT scanner. A scan with our standard protocol for head and neck cancer patients was used (100kV, 80mAs, 5mm slice thickness and backprojection algorithm) and in seven predefined regions (ROI) the signal to noise ratio (SNR) was measured. For the dose optimized protocol, the tube voltage was lowered and the mAs adaptation protocol was used. To improve image quality different percentage of an adaptive statistical iterative reconstruction (ASiR) was applied. For a better resolution we set the slice thickness to 2.5 mm. The mAs adaption range and the percentage of the ASiR reconstruction were varied until we found a combination with the same median SNR in the seven defined ROIs as for our old protocol. For the old and the optimized protocol dose measurements were performed using 25 LiF-TLDs. Organ doses were calculated and the effective dose was determined based on the weighting factors of ICRP103.

The optimized scanning protocol used a voltage of 80kV, a mAs range between 15 and 80, a noise level of 10%, and 50% ASiR reconstruction. The median SNR ratio was slightly better (14% better SNR) with the new protocol. An effective dose of 8 mSv was measured with the original protocol and 4 mSv with the optimized scanning protocol. For organs in the scanning field the dose was reduced by a factor of 2 and outside the field by a factor of 2.2.

Advanced reconstruction algorithms allow a significant dose reduction and an improvement of image resolution, while maintaining the image quality.

Keywords: Perfusion CT imaging, head and neck cancer, effective dose

https://doi.org/10.1515/cdbme-2017-0123

1 Introduction

Perfusion computed tomography (CTP) is a non invasive functional imaging technique to assess the tissue blood circulation, which is associated with tumor angiogenesis. In head and neck tumors CTP imaging has been shown to be a promising tool for treatment prognosis, differentiation of different tumor histologies and differentiation between malignant and benign tumors. [1-4]

The number of CT examinations is continuously increasing, and the average lifetime cancer risk due to medical radiation exposure has been estimated to be as high as 2% [5]. It might be even higher for patients receiving CTP, which consists of repetitive CT scans. Therefore, dose reduction is an important topic, especially for patients with benign tumors and long term survivors.

The current CT technology implements several dose reduction methods, such as automated tube current modulation and iterative reconstruction. These methods have successfully been applied to conventional CT imaging [6] as well as for CTP imaging in the brain region [7]. However their potential to reduce dose for CTP imaging of head and neck cancer has not yet been investigated.

This study was performed to optimize a routinely used CTP protocol for the head and neck region in order to reduce the effective dose to the patient and simultaneously achieve an equal image quality. New iterative reconstruction technology and automated tube current modulation were

---

[*] **Corresponding author: Anja Stüssi[#]:** Department of Radiation Oncology, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland, e-mail: anja.stuessi@usz.ch

Marta Bogowicz[#], Oliver Riesterer, Matthias Guckenberger, Stephanie Tanadini-Lang: Department of Radiation Oncology, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland

Verena Weichselbaumer, Patrick Veit-Haibach: Department of Nuclear Medicine, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland

[#] AS and MB contributed equally

evaluated for dose reduction. The effective dose was estimated based on TLD measurements.

2 Materials and methods

2.1 Scanning protocols and dose optimization

A routinely used protocol suggested by the vendor (Protocol R) was the starting point for our dose optimization process (see **Table 1**). For dose reduction the tube voltage was lowered from 100 to 80 kVp. The milliampere-second (mAs) was adapted based on the attenuation of the body region that was analysed in the scout. The level of adaptation depends on the set noise index, i.e. image noise which influences image quality. We set the noise index to 10%, which was a comparable noise to the images from Protocol R. Such a dose reduction results in a worse image quality. To compensate for that an adaptive statistical iterative reconstruction algorithm (ASiR), provided from GE, was applied. The percentage of ASiR was varied until an optimal setting was found. Additionally, for a better resolution we set the slice thickness to 2.5 mm.

2.2 Image quality measurements

To compare the image quality we measured the signal to noise ratio (SNR) in seven predefined regions (ROI 1-7). The images were registered in the MIMVISTA (MIM Software Inc., Cleveland, OH, USA), which is a registration and contouring tool. The ROIs were drawn as circles of 2 cm diameter, equally distributed over the head and neck region. The ROIs were propagated from CTP of Protocol R to CTP of Protocol O (see **Figure 1**).

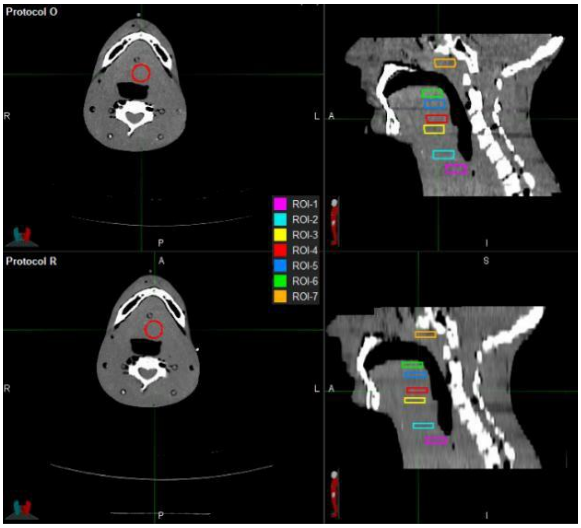


Figure 1: Location of 7 regions of interest (ROI) that were used to evaluate the image quality in the two CTP of protocols: R - routine and O - optimized.

2.3 Dose measurements and effective dose calculation

A tissue equivalent anthropomorphic phantom (Alderson Rando Phantom, Radiology Support Devices, CA, USA) was scanned on a GE Revolution CT scanner (GE Healthcare Technologies, Waukesha, Wisconsin, USA) with both protocols (Protocol R and Protocol O). Dose measurements were performed with LiF-TLD. For each scan, 25 TLDs were distributed over the multiple slices of the phantom inside and outside the scanning field (see **Figure 2**). The allocation of the TLDs was chosen to measure the corresponding organ doses necessary for the effective dose calculation. Additionally, the dose to the eye lenses, a radiation sensitive organ in the scanning field, was measured. The dose to organs in the scanning field was measured using two TLDs, whereas the dose to organs outside of the field was measured with a single TLD. Moreover, for organs located partially in the scanning field (i.e. for brain, oesophagus, skin, bone), the final dose was a result of a weighted average from corresponding TLDs placed in and out of the field.

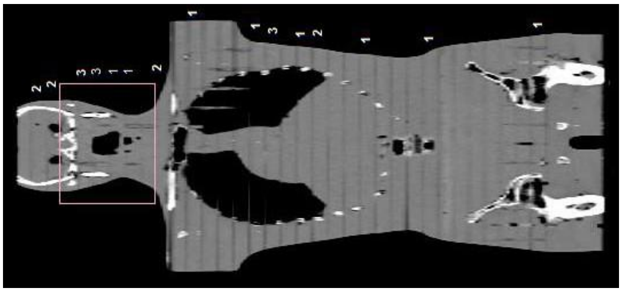


Figure 2: Alderson Phantom with TLDs in the corresponding slice (25 TLDs in total). The red rectangle marks the scanning field with a size of 16 cm.

In the effective dose calculation the weighting factors given by ICRP103 [8] were applied. The fraction of bone marrow distribution inside and outside of the scanning region was estimated based on Cristy et al. [9].

Additionally, the TLD-based effective dose was compared with the effective dose based on the scanner estimation of $CTDI_{vol}$. Dose length product (DLP) was calculated and the effective dose was estimated according to the AAPM 96 [10] using a conversion factor of 0.0031 (for adults for head and neck).

3 Results

The 50% of ASiR was found to provide as good image quality as the Protocol R and consecutively was used in the new optimized protocol (Protocol O) (see **Table 1**).

	Protocol R	**Protocol O**
scan length [cm]	16	16
number of scans	45 (1/s)	45 (1/s)
kV	100	80
mAs	80	----
mAs adaptation [mAs]	----	15-80
noise [%]	----	10
slice thickness [mm]	5	2.5
iterative recon ASiR [%]	0	50

Table 1: CTP parameters for the optimized scanning protocol (Protocol O) and the old in our clinic routinely used protocol (Protocol R).

The median SNR ratio (over all 7 segments) was slightly better (14% better SNR) with the new protocol (see **Figure 3**). ROI 1 shows the largest improvement with 2.5 times better SNR. The Wilcoxon rank sum test showed no significant difference between the two protocols (p>0.1).

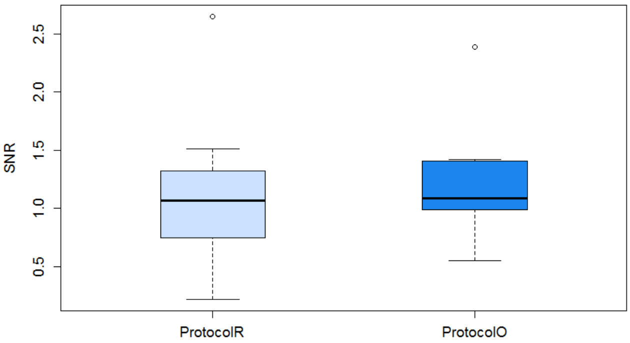


Figure 3: Boxplot of the SNR measurements for the ROI 1 – 7 for both protocols.

The measured organ doses, the effective dose based on the TLD measurements and based on the DLP are shown in **Figure 4** for both protocols (Protocol R, Protocol O). The effective dose decreased with the new protocol. For organs in the scanning field the median dose was reduced by a factor of 2 and outside the field by a factor of 2.2. Additionally, the dose to the eye lenses was reduced by a factor of 1.6. Based on TLD measurements, the Protocol R delivered 8 mSv, whereas the Protocol O delivered only 4 mSv. The effective dose calculated out of DLP was 9 mSv and 5 mSv for the protocol R and protocol O, respectively.

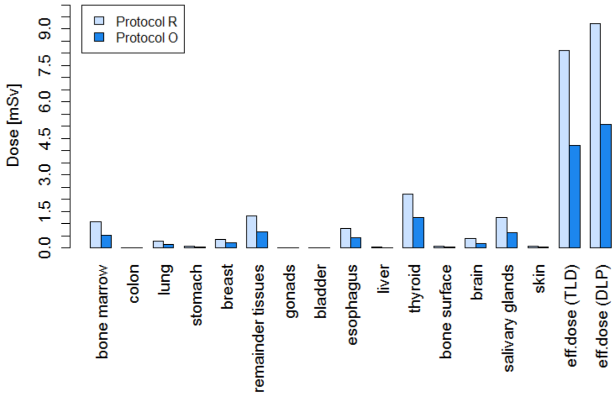


Figure 4: Organ doses, effective doses based on TLD measurements and based on the DLP calculation for protocol R and protocol O.

4 Discussion

In this study we showed that the decreased tube voltage and tube current in combination with iterative reconstruction allowed for an effective dose reduction by a factor of 2 and the preservation of the image quality in the CTP head and neck scan. Additionally, our new protocol improved the out-plane resolution by a factor of 2.

The organ dose comparison showed a dose reduction of 43% to 61% with the new protocol. This is in agreement with the study published by Niesten et al [7], which showed a 50% dose reduction for cerebral brain perfusion imaging. In our study, exceptions were colon and bladder, where the dose was equal. However, these organs were far off the scanning field and the delivered doses were minimal. Generally, for organs in the scanning field the median dose was reduced by a factor of 2 and outside field by a factor of 2.2.

The use of the iterative algorithm ASiR allowed not only the maintenance of image quality but also a slight improvement of the median SNR as well as an improvement in the out-plane resolution. The median SNR decreased by 14% over all the analyzed ROIs. The slice thickness was reduced from 5 mm to 2.5 mm.

The effective dose to the patient was measured using TLD measurements and we additionally calculated the dose based on the DLP given by the scanner. The effective dose estimated from the DLP was slightly higher compared to our TLD measurements. This is in agreement with the study of Bauhs et al [11], who showed that CTDI overestimates the dose in perfusion CT imaging up to 50%, but contradicts studies on conventional CT, which often show that the DLP based method underestimates the effective dose [12]. The TLD measurements enabled us to give an organ specific estimation of the doses.

A limitation of the study is that we assumed an equal image quality in the phantom without contrast agent will result in a similar quality of the calculated perfusion parameters and that no perfusion parameters were calculated in the frame of this study.

5 Conclusion

Iterative reconstruction algorithms such as ASiR allow a significant dose reduction while maintaining the image quality. Therefore CTP protocols should be adapted if iterative reconstruction algorithms are available. The optimized scanning protocol used a voltage of 80kV, a mAs range between 15 and 80 with a noise level of 10%, and 50% ASiR reconstruction.

Author's Statement

Research funding: This study was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The conducted research is not related to either human or animals use

References

[1] Rumboldt Z, Al-Okaili R, Deveikis JP. Perfusion CT for head and neck tumors: pilot study. AJNR 2005; 26:1178–1185

[2] Zima A, Carlos R, Gandhi D, Case I, Teknos T, Mukherji SK. Can pretreatment CT perfusion predict response of advanced squamous cell carcinoma of the upper aerodigestive tract treated with induction chemotherapy? AJNR 2007; 28:328–334

[3] Bisdas S, Baghi M, Wagenblast J, et al. Differentiation of benign and malignant parotid tumors using deconvolution-based perfusion CT imaging: feasibility of the method and initial results. Eur J Radiol 2007; 64:258–265

[4] Nesteruk, M, Lang S., et al. Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma. Radiotherapy and Oncology 2015; 117:125-131

[5] Lee CI, Haims AH, Monico EP, Brink JA, Forman HP. Diagnostic CT scans: assessment of patient, physician, and radiologist awareness of radiation dose and possible risks. Radiology 2004; 231:393 –398

[6] Sagara Y, Hara AK, PavlicekW, Silva AC, Paden RG,Wu Q (2010) Abdominal CT: comparison of low-dose CT with adaptive statistical iterative reconstruction and routine-dose CT with filtered back projection in 53 patients. AJR Am J Roentgenol 195:713–719

[7] Niesten J M, van der Schaaf I C, Riordan A J, de Jong H W, Horsch A D, Eijspaart D, Smit E J, Mali W P and Velthuis B K 2014 Radiation dose reduction in cerebral CT perfusion imaging using iterative reconstruction Eur. Radiol. 24 484–93

[8] ICRP, "The 2007 Recommendations of the International Commission on Radiological Protection. ICRP publication 103," Ann. ICRP 37, 1–332 (2007).

[9] Cristy M., Phys Med Biol. 1981; 26(3):389-400.

[10] AAPM, "The measurement, reporting, and management of radiation dose in CT," AAPM Report No. 96, 2008

[11] Bauhs, J. A., Vrieze, T. J., Primak, A. N., Bruesewitz, M. R., & McCollough, C. H. 2008; CT dosimetry: comparison of measurement techniques and devices 1. *Radiographics*, *28*(1), 245-253.

[12] Geleijns, J., Van Unnik, J. G., Zoetelief, J., Zweers, D., & Broerse, J. J. 1994; Comparison of two methods for assessing patient dose from computed tomography. *The British journal of radiology*, *67*(796), 360-365.

3.3 Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models.

Bogowicz Marta[1], Leijenaar Ralph TH[2], Tanadini-Lang Stephanie[1], Riesterer Oliver[1], Pruschy Martin[1], Studer Gabriela[1], Unkelbach Jan[1], Guckenberger Matthias[1], Konukoglu Ender[3], Lambin Philippe[2]

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[2] Maastricht University Medical Centre - GROW-School for Oncology and Developmental Biology, Department of Radiation Oncology - MAASTRO, Maastricht, The Netherlands.
[3] Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland.

Status of the manuscript:

published Radiother Oncol. 2017, accepted

Author contribution Marta Bogowicz:

- Input to study design
- Image segmentation and registration
- Radiomics analysis at University Hospital Zurich
- Comparison of radiomics results between University Hospital Zurich and MAASTRO clinic
- (Figure 2)
- Training and validation of prognostic models (results for Figures 3-4 and Table 2)
- Preparation of all supplementary files
- Manuscript writing, revision, editing and proofreading

Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models

[1]Marta Bogowicz, [2]Ralph TH Leijenaar, [1]Stephanie Tanadini-Lang, [1]Oliver Riesterer, [1]Martin Pruschy, [1]Gabriela Studer, [1]Jan Unkelbach, [1]Matthias Guckenberger, [3]Ender Konukoglu, [2]Philippe Lambin

[1] Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Rämistrasse 100, 8091 Zürich, Switzerland

[2] Maastricht University Medical Centre - GROW-School for Oncology and Developmental Biology, Department of Radiation Oncology - MAASTRO, Dr. Tanslaan 12, 6229 Maastricht, The Netherlands

[3] Computer Vision Laboratory, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland

Submitted to the Radiotherapy and Oncology on 19th May 2017; accepted on 17th October 2017

Abstract

Purpose

This study investigated an association of post-radiochemotherapy (RCT) PET radiomics with local tumor control in head and neck squamous cell carcinoma (HNSCC) and evaluated the models against two radiomics software implementations.

Materials and methods

649 features, available in two radiomics implementations and based on the same definitions, were extracted from HNSCC primary tumor region in 18F-FDG PET scans 3 months post definitive RCT (training cohort n=128, validation cohort n=50) and compared using the intraclass correlation coefficient (ICC). Local recurrence models were trained, separately for both implementations, using principal component analysis (PCA) and the least absolute shrinkage and selection operator. The reproducibility of the concordance indexes (CI) in univariable Cox regression for features preselected in PCA and the final multivariable models was investigated using respective features from the other implementation.

Results

Only 80 PET radiomic features yielded ICC>0.8 in the comparison between the implementations. The change of implementation caused high variability of CI in the univariable analysis. Both final models performed equally well in the training and validation cohorts (CI>0.7) independent of radiomics implementation.

Conclusion

The two post-RCT PET radiomic models, based on two different software implementations, were prognostic for local tumor control in HNSCC. However, 88% of the features was not reproducible between the implementations.

Keywords: post-radiochemotherapy 18F-FDG PET, radiomics, local tumor control modelling, software implementation, reproducibility

Introduction

Head and neck squamous cell carcinoma (HNSCC) is one of the most common cancers worldwide with tobacco and alcohol consumption as well as HPV infection being the important risk factors. The standard of care for patients with locally advanced HNSCC is definitive radiochemotherapy (RCT). The locoregional recurrence rate is high, exceeding 50% in HPV negative oropharyngeal carcinoma and non-oropharyngeal cancers [1, 2]. A meta-analysis of post-RCT 18F-fluorodeoxyglucose positron emission tomography (18F-FDG PET) studies reported sensitivity and specificity of around 80% in respect to detection of local tumor recurrence or persistence in HNSCC [3]. Additionally, post-RCT FDG PET has been shown to correlate with overall survival [4].

Radiomics, a high throughput method for quantification of medical images, has been shown a promising input for treatment response modelling [5-11]. It is based on a comprehensive and quantitative analysis of a region of interest performed on different levels: shape, intensity, texture and filter-based analysis. Radiomics is a rapidly growing field of research. However, the studies have been predominantly performed in independent single-institution settings and consequently, the importance of workflow standardization has been indicated [5, 6].

Radiomics analysis requires several image pre-processing steps such as region of interest segmentation and extraction as well as image interpolation and discretization. These steps together with image acquisition and reconstruction parameters may influence radiomic features and therefore interchangeability of derived models (i.e. radiomic signatures) [12, 13]. Many institutions use different software packages for the analysis, which are often in-house developed. Although the implementations are based on the same mathematical definitions, it is likely that they will produce different results due to differences in implementation of algorithms as well as pre-processing [13].

To base clinical decisions on a prognostic model, its validation is required [14, 15]. Several strategies, characterized by different strength, can be used. A cross-validation is often implemented as a first step, followed by temporal validation using data from the same institution but from a different period. Finally, to achieve an unbiased validation, an external validation in an independent dataset should be performed [16]. Most of the radiomics studies have used cross-validation to quantify model performance and so far only one model has been validated in an external and independent dataset [17, 18]. Validation is usually performed by the same research group, using the same tools and methodology. However, radiomic features have been shown to vary with image acquisition parameters, pre-processing and contouring [5, 6, 12] and (to our knowledge) none of the previously published studies investigated the reproducibility of a radiomics-based prognostic model in terms of radiomics software implementation.

This study hypothesize that the prognostic value of radiomic features is software implementation dependent. First, we investigated whether the 3 months post-RCT follow-up 18F-FDG PET radiomics is prognostic for tumor recurrence in HNSCC. Two independent models were trained using two independent radiomics implementations and their performance was validated in a separate dataset. Subsequently, the reproducibility of these models was evaluated when their respective radiomic features were calculated with an independent software implementation.

Materials and methods

Imaging protocol and studied population

This retrospective analysis was approved by the local ethical commission. HNSCC patients treated with definitive radiochemotherapy were enrolled in the study (128 patients in the training and 50 patients in the validation cohort). The validation cohort consisted of patients treated in an institutional phase II prospective study (NCT01435252) with a standardized imaging protocol (the same slice thickness and reconstruction algorithm). Surgery or induction chemotherapy were exclusion criteria (biopsy allowed). The characteristic of the studied cohorts is presented in the Table 1. All patients underwent 18F-FDG PET/CT imaging prior to the treatment and 3 months after the end of the treatment as a standard follow-up examination. Depending on patient's body weight, an activity of 170-470 MBq of 18F-FDG was injected intravenously after the measurement of blood sugar level. The PET acquisition was preformed 60 minutes after 18F-FDG injection with a 3 minutes scanning time and 15 cm axial field-of-view at each bed position. Total acquisition time of the PET was 12-18 min. Images were reconstructed with an iterative algorithm (2D or 3D reconstruction in the training cohort and 3D reconstruction in the validation cohort) with an in-plane pixel size and the slice thickness of 2.73 – 5.47 mm and 3.27 – 4.25 mm, respectively. All data was acquired in the same center.

		Training cohort	**Validation cohort**
	Total number of patients	128	50
	Median follow-up (months)	46 (3-156)	16 (3-28)
	Number of local recurrences	38 (30%)	13 (26%)
	T1/T2	43 (34%)	6 (12%)
	T3/T4	85 (66%)	44 (88%)
	Positive	31 (24%)	22 (44%)
	Negative	36 (28%)	28 (66%)
	Unknown	61 (48%)	0
	Oropharynx	91 (71%)	29 (58%)
	Hypopharynx	22 (17%)	7 (14%)
	Larynx	11 (9%)	7 (14%)
	Oral cavity	4 (3%)	7 (14%)
Treatment	**Radiotherapy**	on average 70 Gy (68 – 72 Gy)	70 Gy
	Chemotherapy	Cisplatin (40 mg/m2, up to 7 cycles) or cetuximab (loading dose 400 mg/m2 followed by 250 mg/m2 weekly)	Cisplatin/cetuximab (weekly same doses as in training cohort) with or without consolidation cetuximab (500 mg/m2 biweekly x 6)
PET scanners	**GE Discovery STE**	64 (50%)	39 (78%)
	GE Discovery 690	10 (8%)	6 (12%)
	GE Discovery RX	23 (18%)	5 (10%)
	GE Discovery HR	15 (12%)	
	GE Discovery LS	16 (12%)	

Table 1. Detailed characteristic of studied cohorts.

Image pre-processing and radiomics analysis

Tumors were semi-automatically segmented in the pre-treatment PET scans using a gradient-based method implemented in MIMVISTA (MIM Software Inc., Cleveland, OH, USA). The pre-treatment and post-treatment scans were rigidly registered and contours were transferred to post-treatment scans. To account for differences in image reconstruction grid all scans were rescaled to 5.5 mm cubic voxels using linear interpolation. This corresponds to the smallest resolution in the studied dataset.

The pre-processed images were shared between the institutions. Post-RCT metabolic heterogeneity was studied in the region of the primary tumor (Figure 1). Two independent software implementations were used: implementation from the University Hospital Zurich (USZ) and the MAASTRO clinic (MAASTRO). In total 649 features, which were based on the same definition and available in both implementations, were extracted:

- Shape (n = 8)
- Intensity-based (n = 17)
- Texture: the Gray Level Co-occurrence Matrix (GLCM; n = 24), the Neighborhood Gray Tone Difference Matrix (NGLTDM; n = 4), the Gray-Level Size Zone Matrix (GLSZM; n = 14), the Gray-Level Run Length Matrix (GLRLM; n = 14).
- Filter-based: Wavelet coiflet (n = 568).

The full list of the extracted features is presented in the supplemental material. A bin size of 0.5 SUV was used for image intensity discretization. The consistency of radiomic features calculated in two different implementations was studied using the two-way mixed single measures intraclass correlation coefficient (ICC).

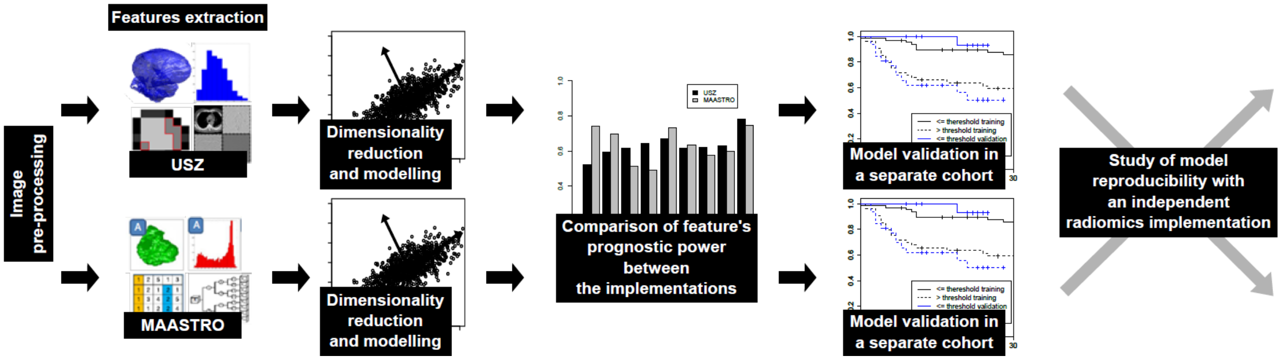


Figure 1. Scheme of the reproducibility analysis of the local tumor control models using two independent radiomics implementation.

Features preselection and comparison of the features' prognostic power between the radiomics implementations

The following feature selection procedure was used. First, a principal component analysis (PCA) was performed to account for inter-feature correlations. The number of retained components was adjusted to represent 95% of data variance. Next, for each principal component one feature was selected to represent it. To that end we determined the feature that correlated the most (the largest Pearson correlation coefficient) with the principal component.

The prognostic power of radiomic features selected in different implementations was investigated in an univariable Cox regression. The models were fitted separately for the USZ and MAASTRO implementations. To quantify the discriminative power of different models the concordance indexes (CI) were calculated and compared between the implementations. The p-value from Cox regression was corrected for the multiple testing using false discovery rate (FDR) < 10% and the number of features defined in the PCA. The statistical analysis was performed in R (v. 3.2.3).

Prediction of local tumor recurrence and model reproducibility between the implementations

To train a final model for the association of the radiomic features derived from post-RCT PET with the likelihood of tumor recurrence, the least absolute shrinkage and selection operator (LASSO) (100 times 5-fold cross-validated) was used for variable selection in multivariable Cox regression. Only the features preselected in the PCA were used in the multivariable analysis. A random sampling with replacement was used to create a different training set in each of the LASSO iterations. In the final model we included only radiomic features with selection rate higher than 70% among all random training sets. Patients were stratified into low- and high-risk of recurrence groups based on a threshold from the receiver operating characteristic curve for local recurrence at 18 months. The threshold was selected to equate the level of sensitivity and specificity. The groups were compared using G-rho test (p-value < 0.05). Two models were trained separately, one on the USZ and one on the MAASTRO feature set. Both models were validated in the independent cohort of patients.

Each trained model, based on the features calculated in one implementation, was later evaluated by calculating its respective features with the other independent implementation (Figure 1). The regression coefficients of the Cox model and the stratification threshold were then fixed. Model performance was quantified using the concordance index (CI). Additionally, the calibration of the models was investigated by calculation of the calibration slope based on the prognostic index [19]. The calibration slope equals 1 evidences the same level of discrimination in the training and validation datasets. Finally, the correlation of hazards obtained with two implementations and the reproducibility of the patients risk group assignments were investigated.

Results

Radiomic features reproducibility between the two implementations

The intraclass correlation coefficient was used to investigate features reproducibility. Out of 649 features, 46 and 80 were characterized by an ICC greater than 0.9 and 0.8, respectively. These were mostly histogram-based (92% of the features in the studied group based on the ICC > 0.8) and texture-based (68%) features calculated on the non-transformed images. The shape features showed intermediate reproducibility (50%), whereas the biggest discrepancy was observed for the wavelet features (supplement Figure 1S). The wavelet features where high-pass filter was applied more than once were the least reproducible. A translation of the 0.5 SUV bin size to the wavelet coefficients was different between the implementations (see supplement section Wavelet). It resulted in a different number of analyzed gray levels in the wavelet maps (supplement Figure 3S). Additionally, the MAASTRO implementation uses an undecimated transform, whereas the USZ implementation uses the decimated one. This influenced the resolution of the analyzed maps.

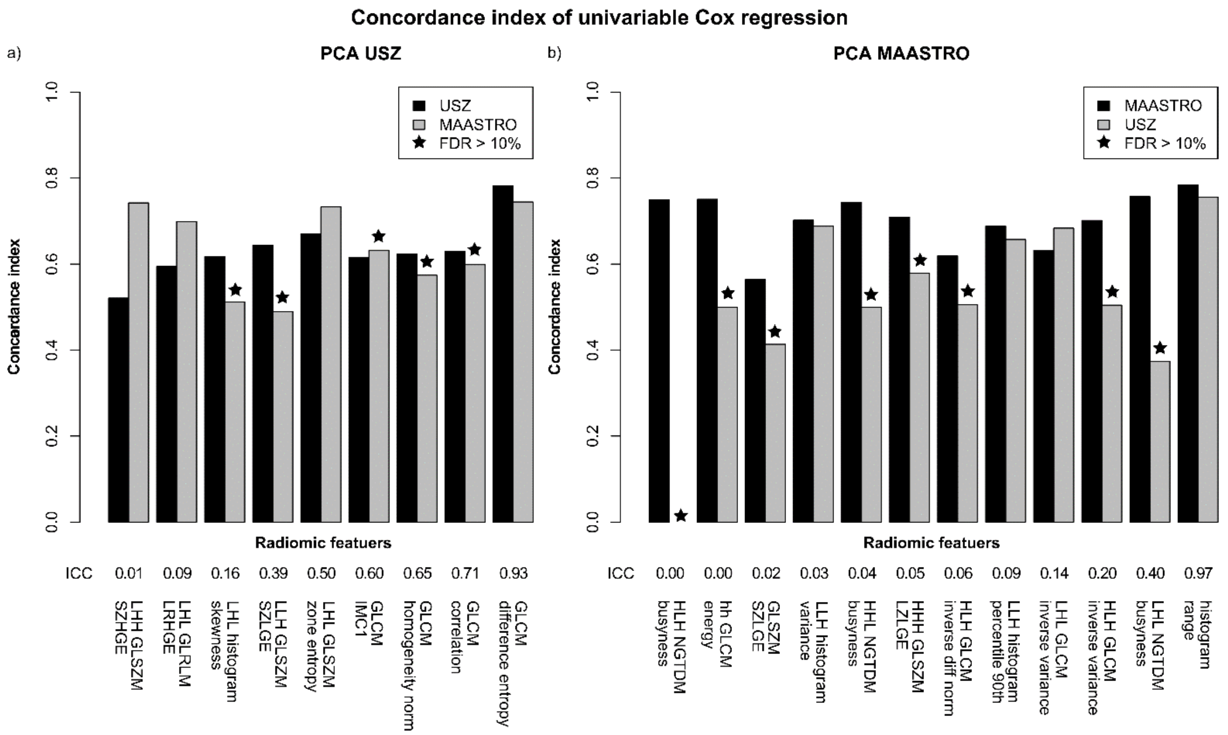


Figure 2. Comparison of the features' prognostic power in the univariable Cox regression between the radiomics implementations. The fit was considered non-significant if false discovery rate (FDR) > 10%. The concordance indexes for the same feature varied between the radiomics implementations and this effect did not depend on the feature's intraclass correlation (ICC) from the implementations comparison. The LLL, LLH, LHL, LHH, HHH, HHL, HLH, HLL – denote the combination of wavelet filters in 3D (L – low-pass, H – high-pass).

Comparison of the features' prognostic power between the radiomics implementations

In the principal component analysis, 31 and 33 components retained the 95% of data variance in the USZ and MAASTRO implementation, respectively. We found only 6 representative features based on the principal components analysis to be the same for both implementations. In a univariable Cox regression, 9 features in USZ and 12 features in MAASTRO implementation yielded a FDR < 10%. Among those features, more than 50% was not significant in the univariable Cox regression when calculated with secondary implementation (Figure 2). Even if the feature was significant in both implementations, a substantial difference in CI was observed. The features were grouped according to their FDR in the secondary implementation (FDR < 10% or FDR ≥ 10%). No significant difference in the ICC values between those groups was observed (Wilcoxon test p-value > 0.05).

Prediction of local tumor recurrence and comparison between the implementations

In the multivariable analysis, GLCM difference entropy was found to be prognostic in the USZ implementation, whereas the histogram range was selected from the MAASTRO implementation. Radiomic features in the final local tumor recurrence models showed high level of reproducibility between the radiomics implementations (ICC > 0.9). A strong correlation (r > 0.9) between GLCM difference entropy and histogram range was observed independent of the implementation. There was a weak, significant correlation between selected radiomic features and tumor volume (r < 0.5).

Both models showed similar prognostic power in the training (5-fold cross-validation) and validation cohorts with CI ranging between 0.70 and 0.76 (Table 2) and allowed for a significant risk group stratification (Figure 3). In the validation cohort, the calibration slope was not significantly different from 1, indicating the preservation of model discriminative power (Table 2). Additionally, the models were prognostic in the group of HPV negative patients (supplement Figure 4S). In both models, tumors with higher risk of recurrence were characterized by a higher post-treatment metabolic heterogeneity (supplement Figure 5S).

The main research question asked in this work was to investigate the model performance when an independent radiomics implementation was used to calculate the hazards. Also in this case, the studied PET radiomics models achieved a very similar performance in terms of the concordance index as well as similar calibration slope (Table 2). It showed that the general discriminative power of the models was not affected by the change of the implementation. On the patient level, a strong correlation was observed between patient rankings based on the features from both implementations (r > 0.9). Most of the patients (around 90%) were correctly classified into low- or high-risk of recurrence group when the independent implementation was used (Figure 4).

		Model developed using radiomic features from	
		MAASTRO	**USZ**
Radiomic features		Histogram range	GLCM difference entropy
Intraclass correlation		0.97	0.93
Concordance Index			
	training	0.76	0.75
	validation	0.73	0.73
	training	0.75	0.74
	validation	0.71	0.72
Calibration slope			
MAASTRO features		1.20 (0.39 – 2.02)*	1.04 (0.27 – 1.95)*
USZ features		1.13 (0.39 – 1.88)*	1.02 (0.20 - 1.83)*

*Table 2. Performance of PET radiomics models for prediction of local tumor control and the stability of radiomic features between two radiomics implementations (USZ and MAASTRO). Underlined values indicates results where the same implementation was used for the training of the model and model performance evaluation, * 95% confidence interval.*

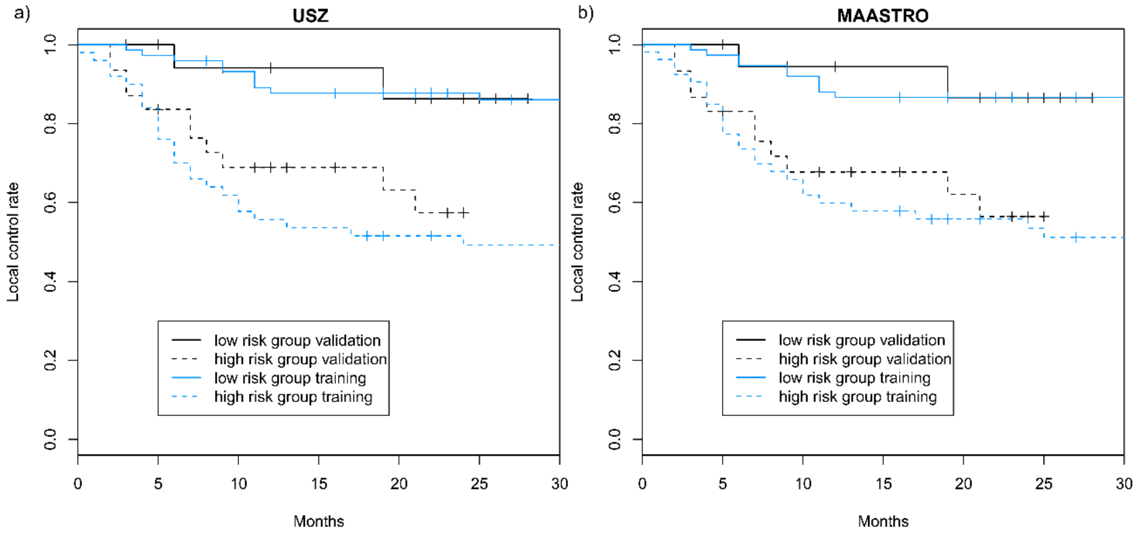


Figure 3. PET radiomics-based local tumor recurrence models: a) USZ implementation, b) MAASTRO implementation. Local control rate curves split significantly (G-rho test p-value < 0.05) in both training and validation cohorts based on the optimal sensitivity-specificity thresholds at 18 months.

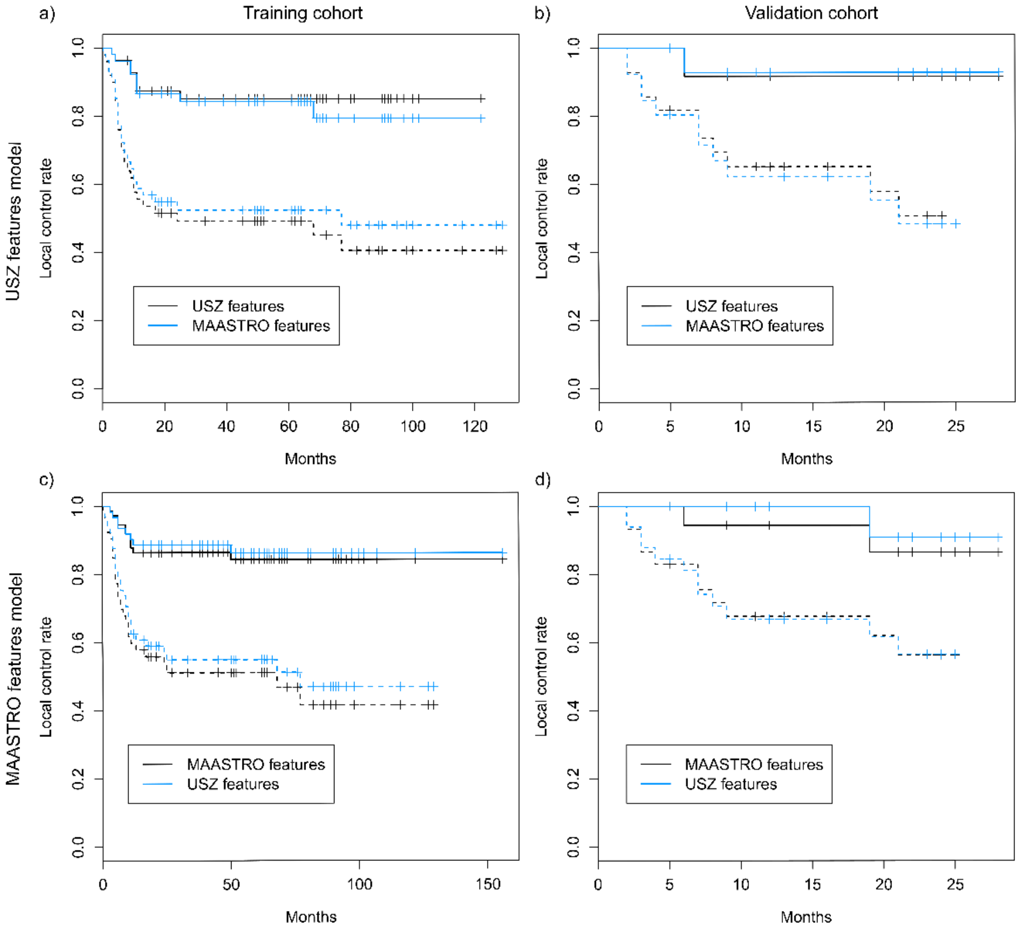


Figure 4. Local control rate curves for low- and high-risk of recurrence groups based on the two PET radiomic models. The curves split significantly independent of the implementation (G-rho test p-value < 0.05).

Discussion

This study investigated the prognostic value of post-RCT PET radiomics in head and neck squamous cell carcinoma and tested the reproducibility of prognostic models between independent radiomics implementations (USZ or MAASTRO). Independent of the radiomics implementation used for model training, the prognostic model for local tumor control showed a good discriminative power with a concordance index higher than 0.7 in both training and validation cohorts. Both models significantly stratified patients into low- and high-risk of recurrence groups. Furthermore, the validation of the models using an independent radiomics implementation resulted in a similar concordance indexes. However, it is important to note that the reproducibility of the models is a consequence of the high ICC between the implementations for the selected features. In the modelling process we have observed that the discriminative power of single radiomic features preselected for the multivariable analysis depended on the radiomics implementation.

The value of post-treatment FDG-PET imaging for assessment of residual disease is currently unclear [3]. Recently, it has been shown in a prospective study that the positive findings on 3 months post-treatment FDG PET are a prognostic factor for overall survival and cancer-specific survival [4]. Additionally, our work shows that the heterogeneity of 3 months post-RCT FDG activity in the region of primary tumor is related to the risk of tumor recurrence. Higher histogram range (range of SUV in the region of primary tumor) and higher GLCM difference entropy corresponded to higher risk of tumor recurrence. We have further shown that these radiomics models can also significantly stratify the HPV negative patients, who belong to a group with a generally bad prognosis. Another study found that a pre-treatment FDG PET radiomics has a similar prognostic power to our post-treatment model [20].

Our prognostic models were trained on a heterogeneous dataset, different PET scanners and reconstruction algorithms were used. However, we were able to validate obtained results on a dataset with a standardized imaging protocol (the same slice thickness and reconstruction algorithm). Our findings should be further validated in datasets from other centers as the lack of calibration between different PET scanners can affect the performance of the models [21]. Additionally, we have defined our region on interest based on the pre-treatment PET images and propagated it to the post-treatment scan. The model reproducibility should be tested against different registration methods for propagation of the delineated tumor volume.

The two radiomics implementations used in this study are based on the same mathematical definition of radiomic features. Additionally, the image pre-processing (image and region of interest resizing) was performed independent of the radiomics implementation and the same bin size was used for image discretization [22]. Nevertheless, a relevant variability in radiomic features value was observed, mostly for the shape and wavelet features. It was most probably caused by differences in mask construction and wavelet transform workflow: especially in the translation of the bin size to the wavelet transformed images and the type of transform (decimated vs undecimated). For more details see supplement sections Contours mapping and Wavelet.. Variations of contour masks constructed from the same DICOM files is also a well-known issue in different treatment planning systems [23]. The comparison of the number of analyzed voxels, as well as minimum, maximum and mean value in the GTV between two radiomics implementations is shown in Figure 2S. The GTV constructed with USZ implementation was always larger then in MAASTRO implementation and consequently the minimum SUV in USZ implementation was always lower. Regarding wavelet transform, the two implementations transferred differently the bin size of 0.5 SUV into the wavelet coefficients space, which resulted in different number of analyzed gray levels (Figure S3). A separate study investigating a discriminative power of wavelet features obtained with the two gray level discretization methods could be conducted to clarify which method is more informative in the context of medical image analysis This study points out differences in radiomics workflow steps, which are rarely described in radiomics studies. Therefore, clear guidelines, such as the Image Biomarker Standardization Initiative [24], providing detailed description of radiomics workflow and implementation are needed. For a workflow comparison purpose, we are also open to share our source code upon request. Nonetheless, we showed that the majority of histogram and texture features was reproducible (i.e. high ICC values) despite the existing differences in contours mask construction. This result suggests that the mask creation is only a minor concern in the standardization aiming for the reproducibility of patients ranking and model prognostic power.

Our final prognostic models for local tumor recurrence were reproducible when the features from the independent radiomics implementation were used, which can be explained by the fact that both models consisted of radiomic features with a high ICC in the comparison between the implementations (ICC > 0.9). Most of the available radiomic features showed a much lower agreement in this comparison. A large variation in concordance indexes was observed for the radiomic features preselected in the principal component analysis. Most of the features preselected in one implementation were not significantly associated with local tumor recurrence in the other implementation in the univariable analysis. This shows, for the first time on a clinically relevant model and dataset, that a model developed by one institution should not be directly transferred to another center, which uses a different radiomics implementation, without rigorous comparison. We recommend that each model, additionally to a detailed description of the radiomics implementation, should be published with a sample dataset and corresponding radiomics signature, such as a recently published digital phantom [25]. This will allow for a comparison of results obtained from a model, before it will be used in a prospective cohort.

In conclusion, this study shows the potential of post-RCT FDG-PET radiomics for early identification of patients with a high risk of local tumor recurrence. It also raises an awareness of the impact of radiomics software implementation on model reproducibility.

Conflict of Interest Statement:

The authors declare that there is no conflict of interest.

Acknowledgements:

The project was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant of the Matching Fund of the University of Zurich. The clinical study used as validation dataset was supported by a research grant from Merck (Schweiz) AG. Additionally, authors acknowledge financial support from the ERC advanced grant (ERC-ADG-2015, n° 694812 - Hypoximmuno), the QuIC-ConCePT project (IMI JU; grant no. 115151). This research is also supported by the Dutch technology Foundation STW (grant n° 10696 DuCAT & n° P14-19 Radiomics STRaTegy), which is the applied science division of NWO, and the Technology Programme of the Ministry of Economic Affairs. Authors also acknowledge financial support from the EU 7th framework program (ARTFORCE - n° 257144, REQUITE - n° 601826), SME Phase 2 (EU proposal 673780 – RAIL), EUROSTARS (DART), the European Program H2020-2015-17 (BD2Decide - PHC30-689715 and ImmunoSABR - n° 733008), Interreg V-A Euregio Meuse-Rhine ("Euradiomics"), Alpe d'HuZes-KWF (DESIGN), Kankeronderzoekfonds Limburg from the Health Foundation Limburg, the Zuyderland-MAASTRO grant and the Dutch Cancer Society.

References

[1] Ang KK, Harris J, Wheeler R et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med, 2010;363:24-35.

[2] Lassen P, Primdahl H, Johansen J. et al Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol. 2014;113:310-316.

[3] Gupta T, Master Z, Kannan S et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging 2011;38:2083-2095.

[4] Kim SA, Roh JL, Kim JS et al. 18F-FDG PET/CT surveillance for the detection of recurrence in patients with head and neck cancer. Eur J Cancer 2016;72:62-70.

[5] Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 2017;44;151-165.

[6] Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61:150-166.

[7] Lambin P, Rios-Velazquez E, Leijenaar R et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441-446.

[8] Bogowicz M, Riesterer O, Ikenberg K, et al. Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma. Int J Rad Onc Biol Phys 2017; In press.

[9] Lambin P, Leijenaar RTH, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017; In press.

[10] Coroller TP, Grossmann P, Hou Y et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol. 2015;114:345-50.

[11] van Timmeren JE, Leijenaar RTH, van Elmpt W et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol. 2017;123:363-369

[12] Larue RT, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. Br J Radiol 2017;90:20160665.

[13] Court LE, Fave X, Mackin D, Lee J, Yang J, Zhang L. Computational resources for radiomics. Translational Cancer Research 2016;5.

[14] Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Cancer 2015;112:251-259.

[15] Rios Velazquez E, Hoebers F, Aerts HJ et al. Externally validated HPV-based prognostic nomogram for oropharyngeal carcinoma patients yields more accurate predictions than TNM staging. Radiother Oncol. 2014;113:324-330.

[16] Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338.

[17] Leijenaar RT, Carvalho S, Hoebers FJ et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015;54:1423-1429.

[18] Aerts HJWL, Rios Velazquez E, Leijenaar RTH et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

[19] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Medical Research Methodology, 2013;13.

[20] Bogowicz M, Riesterer O, Stark LS et al. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol. 2017; In press

[21] Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. AJR Am J Roentgenol. 2010;195:310-320.

[22] Leijenaar R T H, Nalbantov G, Carvalho S et al. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep. 2015;5.

[23] Kulkarni BS, Sharma SD, Hansenet V et al. A prospective study of OAR volume variations between two different treatment planning systems in radiotherapy. International Journal of Cancer Therapy and Oncology 2015;3.

[24] Zwanenburg A, Leger S, Vallières M, Löck S, for the Image Biomarker Standardisation Initiative: Image Biomarker Standardisation Initiative. arXiv:161207003v4 2017.

[25] Lambin P. Radiomics Digital Phantom. Cancer Data 2016, DOI:10.17195/candat.2016.08.1.

Supplement

Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models

[1]Marta Bogowicz, [2]Ralph TH Leijenaar, [1]Stephanie Tanadini-Lang, [1]Oliver Riesterer, [1]Martin Pruschy, [1]Gabriela Studer, [1]Jan Unkelbach, [1]Matthias Guckenberger, [3]Ender Konukoglu, [2]Philippe Lambin

[1] Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Zurich, Switzerland

[2] Maastricht University Medical Centre - GROW-School for Oncology and Developmental Biology, Department of Radiation Oncology - MAASTRO, Maastricht, The Netherlands

[3] Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland

Radiomics features

The following approaches were used to determine radiomics: shape analysis, histogram of intensities analysis, texture analysis (the Gray Level Co-occurrence Matrix, the Neighborhood Gray Tone Difference Matrix, the Gray Level Size Zone Matrix and the Gray Level Run Length Matrix) and wavelet transform analysis.

Contours mapping

Both radiomics implementations are fully DICOM compatible. The contours used in the analysis were saved in the DICOM format. In the USZ implementation, the contour points were translated to the voxel coordinates of the analyzed image. Next, those voxels were used to find a contour polygon and points inside this polygon using the pointPolygonTest function from Python cv2 library (version 2.4.6). In the MAASTRO implementation an algorithm, which is equivalent to the standard mask construction algorithm available in Computational Environment for Radiotherapy Research (CERR), was used [1]. The contour points were used directly to construct contour polygon. Next, the image-specific coordinates of voxels inside this polygon were found.

The difference in mask construction resulted in generally larger masks in USZ implementation in comparison to MAASTRO implementation (see Figure 2S).

Discretization of image intensities

Both radiomics implementations used in this study the fixed bin size discretization of image intensities, with the bin size of 0.5 SUV. The lower range of bin 1 was defined as the minimum value in the analyzed region of interest. However, two different discretization schemes were used in the transformed images (see the Wavelet transform section).

Shape

In the USZ implementation: to calculate shape features contours were transformed onto 1 mm isotropic grid; the volume and surface estimation was done using marching cubes algorithm implemented in the VTK library [2]. In the MAASTRO implementation: shape features were calculated from the same resolution images as the texture features; surface area was calculated using the isosurface triangulation and volume was defined as the voxel's volume times number of voxels in the region of interest.

1) V – volume
2) A – surface
3) $compactness\ 1 = \frac{V}{(\pi^2 A)^{\frac{3}{2}}}$
4) $compactness\ 2 = 36\pi \frac{V^2}{A^3}$
5) $spherical\ disproportion = \frac{A}{4\pi R^2}$

where: R is the radius of a sphere with the same volume as the tumor.

6) $sphericity = \frac{(36\pi V^2)^{\frac{1}{3}}}{A}$
7) $surface\ to\ volume\ ratio = \frac{A}{V}$
8) $maximum\ 3D\ diameter$ - the largest pairwise Euclidian distance between voxels of the region of interest

Histogram of intensities

The parameters from the histogram of intensities were calculated on PET images before a discretization. Let X denotes the intensities of the 3D image with N voxels. $\bar{X}$ - mean of X, N_g - number of gray levels in the image, p_i - the occurrence probability of gray level i.

9) $mean = \frac{1}{N}\sum_{i=1}^{N} X_i$
10) $standard\ devation = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$
11) $variance = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$
12) $skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}\right)^3}$
13) $mean\ absolut\ deviation = \frac{1}{N}\sum_{i=1}^{N}|X_i - \bar{X}|$
14) $robust\ mean\ absolut\ deviation = \frac{1}{N}\sum_{i=1}^{N_{10-90}}\left|X_{10-90,i} - \bar{X}_{10-90}\right|$

where: N_{10-90}- number of voxel in the range from and 10th percentile and 90th percentile, $\bar{X}_{10-90}$ - mean value of voxel in the range from 10th percentile and 90th percentile

15) $energy = \sum_{i=1}^{N} X_i^2$
16) $entropy = -\sum_{i=1}^{N_g} p_i \log_2 p_i$
17) $root\ mean\ square = \sqrt{\frac{energy}{N}}$
18) $uniformity = \sum_{i=1}^{N_g} p_i$
19) $minimum\ value = \min(X)$
20) $maximum\ value = \max(X)$
21) $median$ – the median value of X
22) $10^{th} percentile$ – 10th percentile of X
23) $90^{th} percentile$ – 90th percentile of X
24) interqurtile range = $90^{th} percentile - 10^{th} percentile$
25) $range = \max(X) - \min(X)$

Texture

The Gray Level Co-occurrence Matrix

The parameters from the Gray Level Co-occurrence Matrix [3] were calculated in all 26 directions with a distance of one voxel. The final parameters were the average of all directions. If one of the voxels had a 'not a number' value the pair was not taken into account in the calculations. Let P_{ij} denotes the (i,j) entry in the Gray Level Co-occurrence Matrix, N_g - number of gray tones in a studied structure, $P_{xi} = \sum_{j=1}^{N_g} P_{ij}$, $P_{yj} = \sum_{i=1}^{N_g} P_{ij}$, $P_{x+y}(k) = \sum_{j=1}^{N_g} \sum_{i=1}^{N_g} P_{ij}$, where $k = i + j$, $P_{x-y}(k) = \sum_{j=1}^{N_g} \sum_{i=1}^{N_g} P_{ij}$, where $k = |i - j|$.

26) $energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}^2$
27) $contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 P_{ij}$
28) $correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$

where: $\mu_x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} iP_{ij}$, $\mu_y = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} jP_{ij}$, $\sigma_x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}(i - \mu_x)^2$, $\sigma_y = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}\left(j - \mu_y\right)^2$

29) $sum\ of\ squares = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P_{ij}$

where μ – mean of P

30) $inverese\ difference\ moment\ (homogeneity) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{1+(i-j)^2}$

31) $inverese\ difference\ moment\ normalized = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{1+\frac{(i-j)^2}{N_g^2}}$

32) $inverese\ difference\ = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{1+|i-j|}$

33) $inverese\ difference\ normalized = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{1+\frac{|i-j|}{N_g}}$

34) $sum\ of\ average = \sum_{i=2}^{2N_g} i \cdot\ P_{x+y}(i)$

35) $sum\ entropy = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log_2 P_{x+y}(i)$

36) $entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij}$

37) $difference\ entropy = -\sum_{i=0}^{N_g-1} P_{x-y}(i)\log_2 P_{x-y}(i)$

38) $information\ measure\ of\ correlation\ 1 = \frac{-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij} - \left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right)}{max\left\{\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{xi}\log_2 P_{xi}\right),\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{yj}\log_2 P_{yj}\right)\right\}}$

39) $information\ measure\ of\ correlation\ 2 =$

$\sqrt{1-\exp\left[-2-\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right)+\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij}\right]}$

40) $maximal\ correlation\ coefficient = \sqrt{\text{second largest eigenvalue of } \sum_{k=1}^{N_g}\frac{P_{ik}P_{jk}}{P_{xi}P_{ji}}}$

41) $joint\ maximum = \max(P_{ij})$

42) $joint\ average = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} iP_{ij}$

43) $difference\ average = \sum_{i=0}^{N_g-1} iP_{x-y}(i)$

44) $dissimilarity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|\,P_{ij}$

45) $inverese\ variance = 2\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{(i-j)^2}$

46) $autocorrelation = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP_{ij}$

47) $cluster\ tendency = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(i+j-\mu_x-\mu_y\right)^2 P_{ij}$

48) $cluster\ shade = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(i+j-\mu_x-\mu_y\right)^3 P_{ij}$

49) $cluster\ prominence = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left(i+j-\mu_x-\mu_y\right)^4 P_{ij}$

The Neighborhood Gray Tone Difference Matrix

The Neighborhood Gray Tone Difference Matrix [4] was calculated based on 26 adjacent voxels. The voxels with ‘not a number’ value were excluded from the average over the neighborhood region. Let s_i denotes the i^{th} entry in the Neighborhood Gray Tone Difference Matrix, N_i - the number of voxels having gray tone i, G - number of gray tones in a studied structure, n - number of studied voxels.

50) $coarseness = \left[\epsilon + \sum_{i=1}^{G} \frac{N_i}{n} s_i\right]^{-1}$

where ϵ is a small number to prevent coarseness becoming infinite.

51) $contrast = \left[\frac{1}{N_g(N_g-1)} \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{N_i}{n} \frac{N_j}{n} (i-j)^2\right] \left[\frac{1}{n} \sum_{i=1}^{G} s_i\right]$

where N_g is the total number of different gray levels present in the image.

52) $busyness = \frac{\sum_{i=1}^{G} \frac{N_i}{n} s_i}{\sum_{i=1}^{G} \sum_{j=1}^{G} i\frac{N_i}{n} - j\frac{N_j}{n}}$

for $\frac{N_i}{n} \neq 0$ and $\frac{N_j}{n} \neq 0$

53) $complexity = \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{|i-j|\left(\frac{N_i}{n} s_i + \frac{N_j}{n} s_j\right)}{N_i + N_j}$

for $N_i \neq 0$ and $N_j \neq 0$

The Gray Level Size Zone Matrix

In the Gray Level Size Zone Matrix [5] calculation the voxels with ‘not a number’ value were excluded from the analysis. Let P_{ij} denotes the (i,j) entry in the Gray Level Size Zone Matrix, i - gray value, j - size, n_r - number of homogeneous areas inside a studied structure and $p_{ij} = P_{ij}/n_r$, $\mu_i = \sum_{i=1}^{M} \sum_{j=1}^{N} iP_{ij}$, $\mu_j = \sum_{i=1}^{M} \sum_{j=1}^{N} iP_{ij}$.

54) $gray\ level\ non-uniformity = \frac{1}{n_r} \sum_{i=1}^{M} \left(\sum_{j=1}^{N} P_{ij}\right)^2$

55) $size\ zone\ non-uniformity = \frac{1}{n_r} \sum_{j=1}^{N} \left(\sum_{i=1}^{M} P_{ij}\right)^2$

56) $small\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij}}{j^2}$

57) $large\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij} \cdot j^2$

58) $low\ gray\ level\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij}}{i^2}$

59) $high\ gray\ level\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij} \cdot i^2$

60) $small\ zone\ low\ gray\ level\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij}}{i^2 \cdot j^2}$

61) $small\ zone\ high\ gray\ level\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij} \cdot i^2}{j^2}$

62) $large\ zone\ low\ gray\ level\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij} \cdot j^2}{i^2}$

63) $large\ zone\ high\ gray\ level\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij} \cdot i^2 \cdot j^2$

64) $size\ precentage = \frac{n_r}{number\ of\ voxels\ in\ the\ studied\ structure}$

65) $gray\ level\ variance = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} (i - \mu_i)^2 P_{ij}$

66) $size\ zone\ variance = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(j - \mu_j\right)^2 P_{ij}$

67) $zone\ entropy = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\log(P_{ij})$

The Gray Level Run Length Matrix

In the Gray Level Run Length Matrix [6] calculation the voxels with ‘not a number’ value were excluded from the analysis. Let P_{ij} denotes the (i, j) entry in the Gray Level Run Length Matrix, i - gray value, j - size, n_r - number of homogeneous areas inside a studied structure and $p_{ij} = P_{ij}/n_r$, $\mu_i = \sum_{i=1}^{M}\sum_{j=1}^{N} iP_{ij}$, $\mu_j = \sum_{i=1}^{M}\sum_{j=1}^{N} iP_{ij}$.

68) $gray\ level\ non - uniformity = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} P_{ij}\right)^2$

69) $run\ length\ non - uniformity = \frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M} P_{ij}\right)^2$

70) $short\ length\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{j^2}$

71) $long\ length\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot j^2$

72) $low\ gray\ level\ length\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2}$

73) $high\ gray\ level\ length\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2$

74) $short\ length\ low\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2\cdot j^2}$

75) $short\ length\ high\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot i^2}{j^2}$

76) $long\ length\ low\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot j^2}{i^2}$

77) $long\ length\ high\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2\cdot j^2$

78) $length\ precentage = \frac{n_r}{number\ of\ voxels\ in\ the\ studied\ structure}$

79) $gray\ level\ variance = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}(i - \mu_i)^2 P_{ij}$

80) $run\ length\ variance = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(j - \mu_j\right)^2 P_{ij}$

81) $run\ entropy = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\log(P_{ij})$

Wavelet

The coiflet wavelet of 1st order was used in the wavelet transform in both implementations [7].The decimated wavelet transform (with downsmaple factor of 2) was used by the USZ implementation, whereas the MAASTRO implemenatition used the undecimated one. It influenced resoltuion of analyzed images. In the MAASTRO implementation no contour adaptation was needed, because the wavelet decompositions are of the same resolution as the original image. In the USZ implementation contours had to be resampled to the lower resolution grid.

The two implementations differed also in terms of wavelet maps discretization. In the USZ implementation, the sum of energy in the wavelet maps was normalized to the energy of the original image. Next, the same bin size of 0.5 was used. In the MAASTRO implementation, the bin size in each of the wavelet maps was adapted to match the number of bins analyzed in the original image. In both cases the minimum value in a single wavelet map corresponded to the lower range of bin number 1.

Results

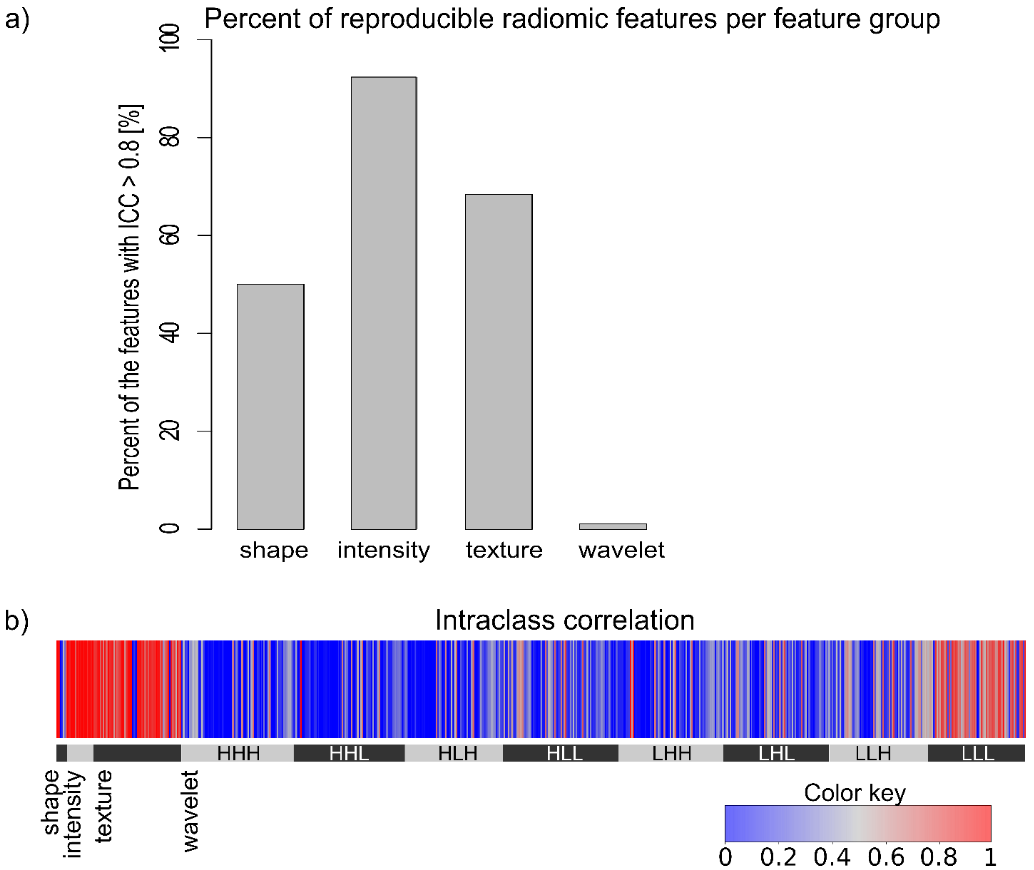


Figure 1S. The reproducibly of radiomics features between the implementations. a) Histogram of reproducible radiomic features. The intensity and texture features showed a high level of agreement between the implementations. The irreproducibility of shape and wavelet features was caused by differences in mask extraction and wavelet maps normalization. b) Intraclass correlation coefficient for different features, H – high-pass filet, L – low-pass filter. The features where high-pass filter was applied more than once were the least reproducible. The difference between the implementation in wavelet maps normalization did not have a big influence on the low-pass filtered images.

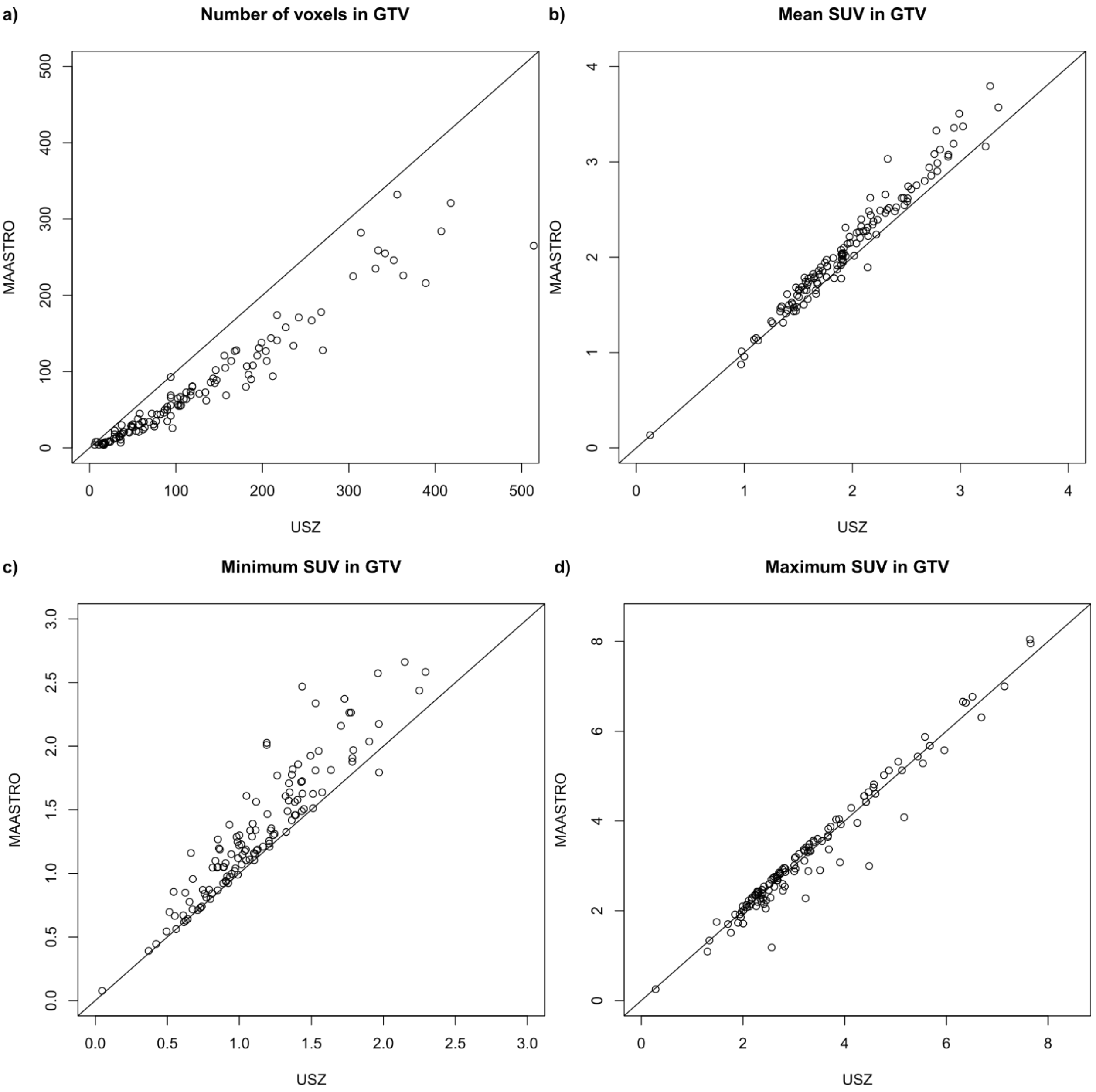


Figure 2S. The comparison of mask extraction algorithms used by the different radiomics implementation. Each point corresponds to one patient in the training cohort. The masks extracted with USZ implementation were generally larger (larger number of analyzed voxels (a)). It also resulted in a lower minimum SUV (c) observed in USZ implementation, whereas the maximum (d) and mean SUV (b) were less affected. A 1:1 line was plotted for the comparison.

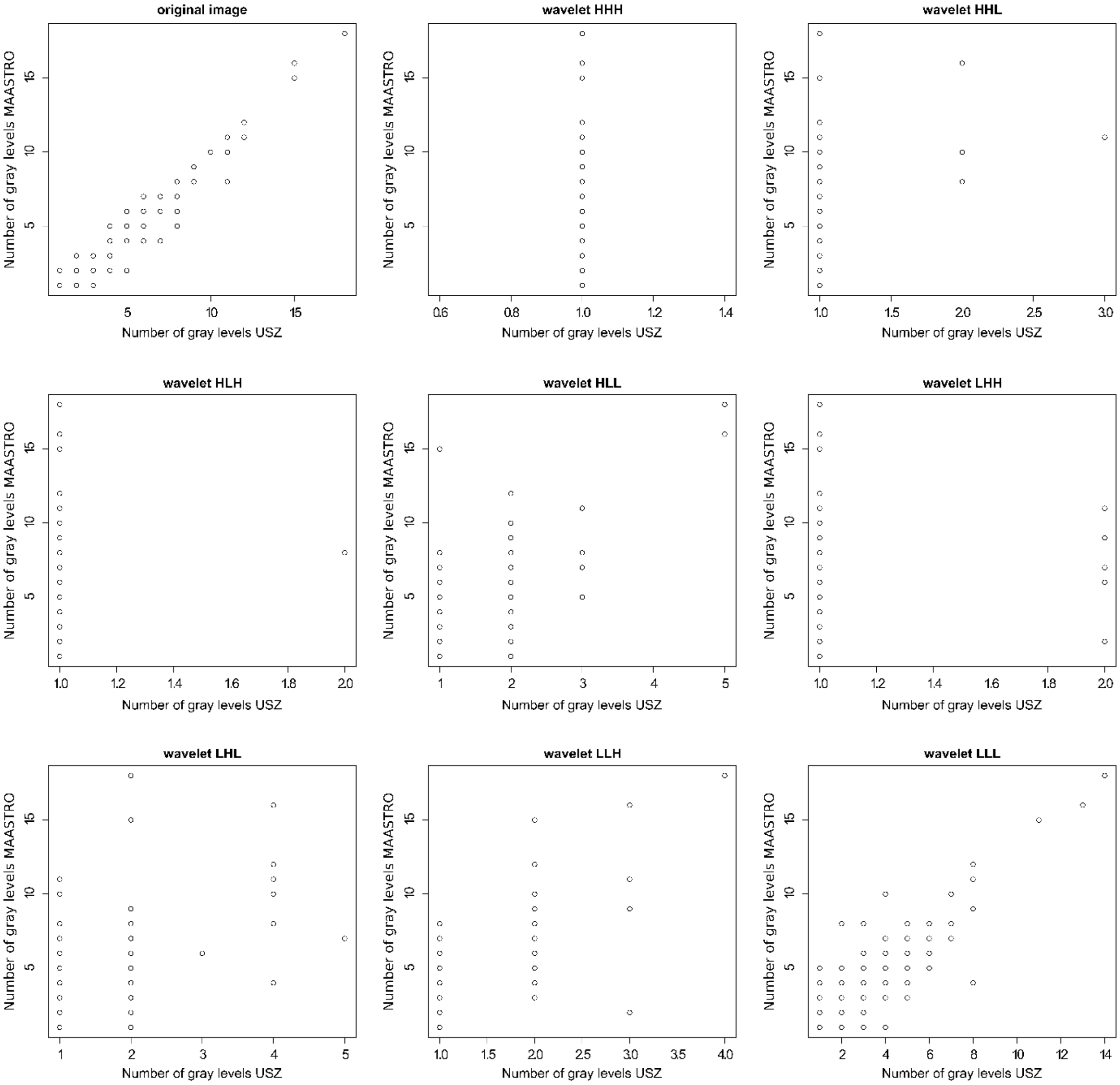


Figure 3S. The comparison of the number of gray level used in the analysis by the two independent implementations (USZ and MAASTRO). Each point corresponds to one patient in the training cohort. Although, both implementations used the same bin size (0.5 SUV) its translation to the bin size in the wavelet transformed maps differed, resulting in different number of gray levels.

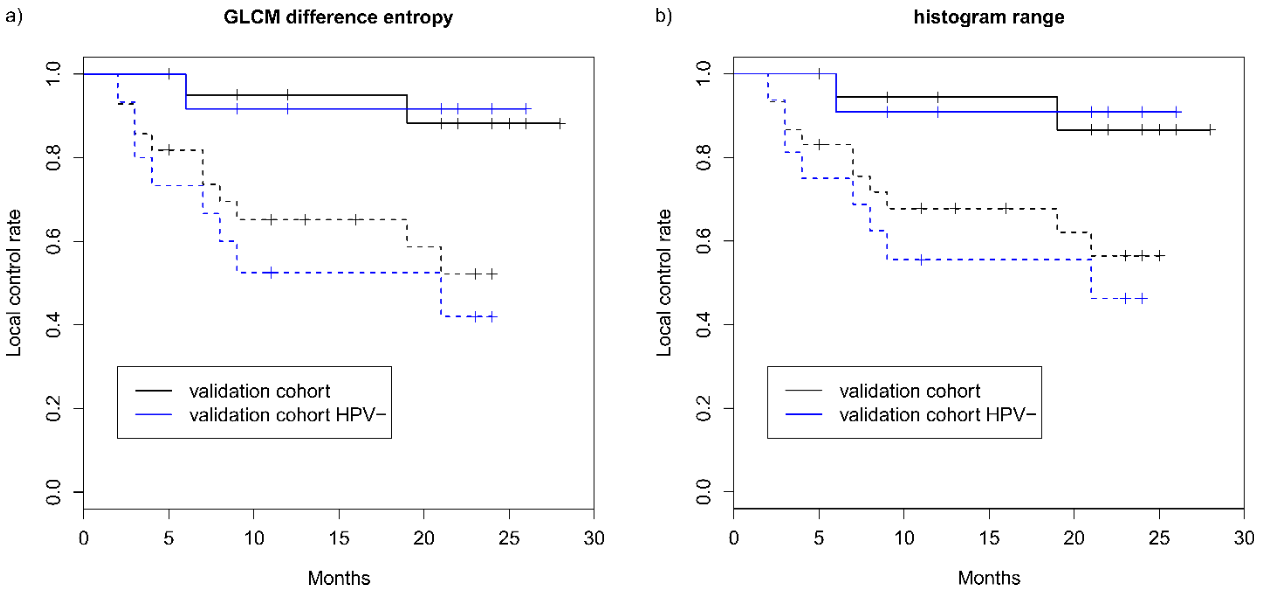


Figure 4S. PET radiomics-based local tumor recurrence models: a) USZ implementation, b) MAASTRO implementation. The models are prognostic not only in the validation cohort but also in the subgroup of HPV negative patients (CI_{USZ} = 0.78, $CI_{MAASTRO}$ = 0.82). The local control curves split significantly (G-rho test p-value < 0.05) based on the optimal sensitivity-specificity thresholds at 18 months.

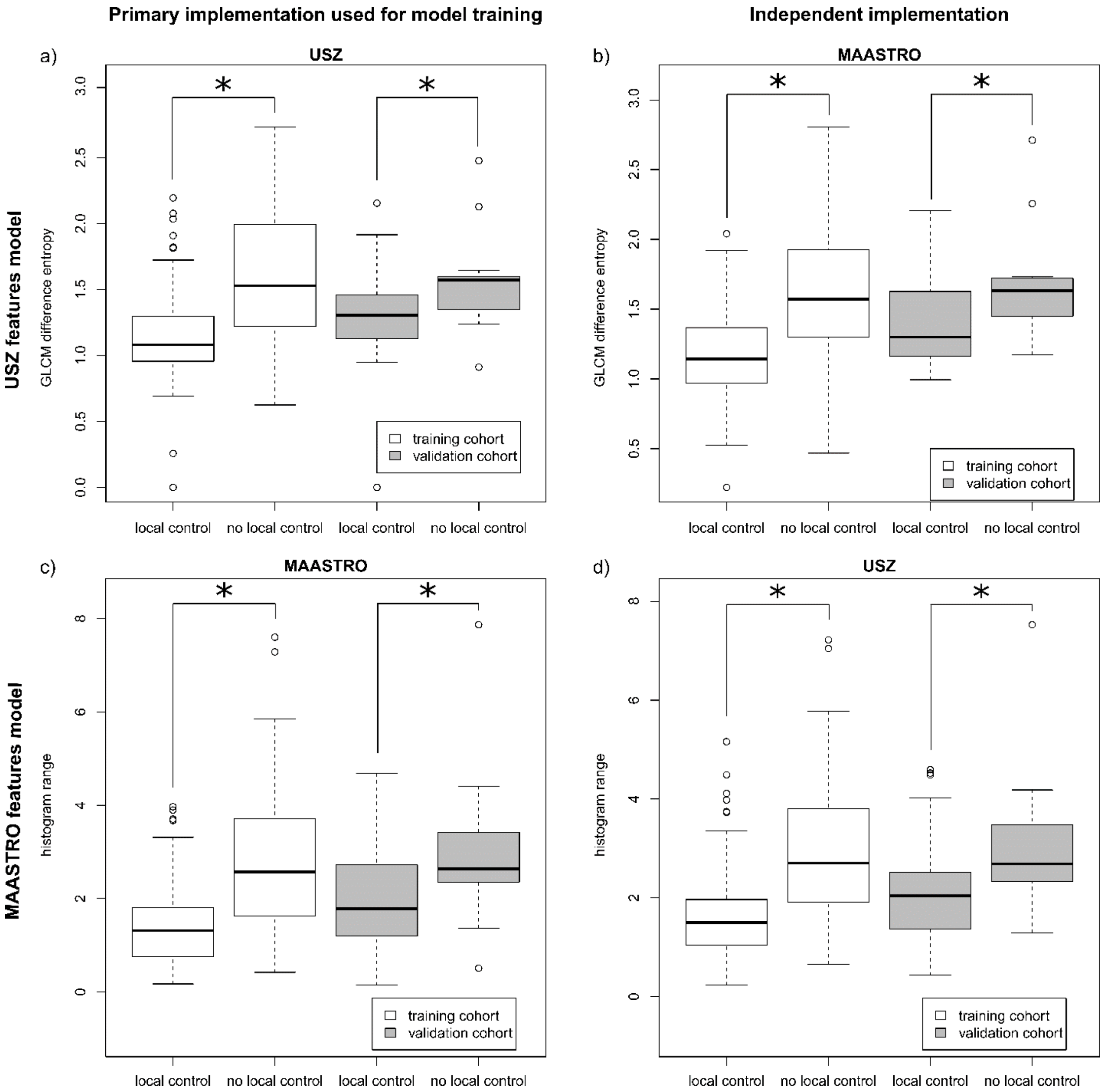


Figure 5S. Post-RCT PET radiomics signature prognostic for local tumor recurrence. In both implementations radiomic features selected in the final models were significantly different for patients with controlled tumors and with recurrences (* Wilcoxon rank-sum test $p < 0.05$).

References

[1] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. Med Phys 2003;30: 979-985

[2] Schroeder W, Martin K, Lorensen B. The Visualization Toolkit (4th ed.). Kitware, 2006.

[3] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classiification. IEEE Trans Syst Man Cybern. 1973;3:610-621.

[4] Amadasun M, King R. Texural Features Corresponding to Textural Properties. IEEE Trans Syst Man Cybern. 1989;19.

[5] Thibault G, Fertil B, Navarro C, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. Pattern Recognition Inf Process. 2009;:140-145.

[6] Galloway MM. Texture analysis using gray level run lengths. Computer Graphics and Image Processing 1975;4:172-197.

[7] Beylkin G, Coifman R, Rokhlin V, Fast wavelet transforms and numerical algorithms, Comm. Pure Appl. Math. 1991;44:141-183

3.4 Stability of radiomic features in CT perfusion maps.

Bogowicz Marta[1], Riesterer Oliver[1], Bundschuh Ralph A[2], Veit-Haibach Patrick[3,4], Hüllner Martin[3,4], Studer Gabriela[1], Stieb Sonja[1], Glatz Stefan[1], Pruschy Martin[1], Guckenberger Matthias[1], Tanadini-Lang Stephanie[1].

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.

[2] Department of Nuclear Medicine, University Hospital Bonn, Bonn, Germany.

[3] Department of Nuclear Medicine, University Hospital Zurich and University of Zurich, Zurich Switzerland.

[4] Department of Diagnostic and Interventional Radiology, University Hospital Zurich and University of Zurich, Zurich Switzerland.

Status of the manuscript:

published in Phys Med Biol. 2016;61(24):8736-8749.

Author contribution Marta Bogowicz:

- Development of a DICOM-compatible software implementation for radiomics calculation
- Computation of perfusion maps using different settings, computation of radiomics
- Full stability analysis and interpretation (results for Figures 2-4 and Table 3)
- Preparation of all supplementary files
- Manuscript writing, revision, editing and proofreading.

IOP Publishing | Institute of Physics and Engineering in Medicine Physics in Medicine & Biology

Phys. Med. Biol. 61 (2016) 8736–8749 doi:10.1088/1361-6560/61/24/8736

Stability of radiomic features in CT perfusion maps

M Bogowicz[1], **O Riesterer**[1], **R A Bundschuh**[4], **P Veit-Haibach**[2,3], **M Hüllner**[2], **G Studer**[1], **S Stieb**[1], **S Glatz**[1], **M Pruschy**[1], **M Guckenberger**[1] **and S Tanadini-Lang**[1]

[1] Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Rämistrasse 100, 8091 Zürich, Switzerland
[2] Department of Nuclear Medicine, University Hospital Zurich, University of Zurich, Rämistrasse 100, 8091 Zürich, Switzerland
[3] Department of Diagnostic and Interventional Radiology, University Hospital Zurich, University of Zurich, Rämistrasse 100, 8091 Zürich, Switzerland
[4] Department of Nuclear Medicine, University Hospital Bonn, Sigmund-Freud-Straße 25, D-53127 Bonn, Germany

E-mail: marta.nesteruk@usz.ch

Received 26 July 2016, revised 25 October 2016
Accepted for publication 3 November 2016
Published 28 November 2016

Abstract

This study aimed to identify a set of stable radiomic parameters in CT perfusion (CTP) maps with respect to CTP calculation factors and image discretization, as an input for future prognostic models for local tumor response to chemo-radiotherapy.

Pre-treatment CTP images of eleven patients with oropharyngeal carcinoma and eleven patients with non-small cell lung cancer (NSCLC) were analyzed. 315 radiomic parameters were studied per perfusion map (blood volume, blood flow and mean transit time). Radiomics robustness was investigated regarding the potentially standardizable (image discretization method, Hounsfield unit (HU) threshold, voxel size and temporal resolution) and non-standardizable (artery contouring and noise threshold) perfusion calculation factors using the intraclass correlation (ICC). To gain added value for our model radiomic parameters correlated with tumor volume, a well-known predictive factor for local tumor response to chemo-radiotherapy, were excluded from the analysis. The remaining stable radiomic parameters were grouped according to inter-parameter Spearman correlations and for each group the parameter with the highest ICC was included in the final set. The acceptance level was 0.9 and 0.7 for the ICC and correlation, respectively.

The image discretization method using fixed number of bins or fixed intervals gave a similar number of stable radiomic parameters (around 40%). The potentially standardizable factors introduced more variability into radiomic parameters than the non-standardizable ones with 56–98% and 43–58% instability rates, respectively. The highest variability was observed for voxel size (instability rate >97% for both patient cohorts). Without standardization of CTP calculation factors none of the studied radiomic parameters were stable. After standardization with respect to non-standardizable factors ten radiomic parameters were stable for both patient cohorts after correction for inter-parameter correlations.

Voxel size, image discretization, HU threshold and temporal resolution have to be standardized to build a reliable predictive model based on CTP radiomics analysis.

Keywords: radiomics, CT perfusion, radiomic features stability

S Online supplementary data available from stacks.iop.org/PMB/61/8736/mmedia

Introduction

Cancer is a heterogeneous disease in regard to etiology, pathogenesis, therapy response and prognosis (Lamlertthon *et al* 2011). Tumor response to therapy varies not only among patients but also within the tumor itself. For optimizing treatment strategies, identification of biomarkers may be essential. Imaging biomarkers are of special interest as they provide spatial information on tumor biology and are acquired non-invasively.

Tumor perfusion assessed by computed tomography perfusion (CTP) is a surrogate marker for tumor angiogenesis (Goh *et al* 2008, Ling *et al* 2014, Luczynska *et al* 2014). Based on CTP images, blood in the vasculature can be described by three main parameters: blood volume (BV), blood flow (BF) and mean transit time (MTT) (Lee *et al* 2003). Tumors or tumor sub-volumes characterized by poor angiogenesis are likely to be hypoxic and thus are more resistant to radiation therapy (Vaupel and Mayer 2007). Several studies showed that the perfusion inside a tumor prior to surgery or radiation therapy is correlated with local tumor control (Preda *et al* 2014, Yeung *et al* 2015).

In recent years, radiomics has become increasingly important for medical image characterization, both in terms of volume segmentation (Kumar and Mehta 2011, Lin *et al* 2016) and prediction of treatment response (Aerts *et al* 2014, Itakura *et al* 2015, Parmar *et al* 2015, Yang *et al* 2015). Radiomics uses mathematical definitions to investigate regularity (or periodicity), directionality and complexity, which are the prominent perceptual texture characteristics (Rao and Lohse 1993). Altogether radiomic features provide much more information about a region of interest than the mean or maximum intensity values generally used in clinical medicine. Various approaches have been developed to identify texture in an image: the statistical-based approach, the model-based approach and the transform-based approach. In the statistical-based approach, the texture is described by non-deterministic properties that influence the distributions of the gray levels of an image (Materka 2004). First-order statistical parameters are calculated using a histogram of intensities in the image and do not contain information about the spatial distribution of pixels. Higher-order statistical parameters are computed based on the matrices describing the spatial changes in pixel intensities, such as the gray level co-occurrence matrix (Haralick *et al* 1973). Another texture analysis method is the model-based approach, which uses the stochastic or fractal models to interpret the structure (Materka 2004). Alternatively, the transform-based approach allows analyzing the image in the frequency domain. A wavelet transform is of special interest as it provides a wide range of choices for the wavelet function and preserves the spatial localization of the frequencies (Materka 2004).

Further improvement of prognostic models based on CTP using radiomics analysis is an interesting but not yet studied topic. To develop a prognostic model for tumor response based on radiomic parameters, a set of stable radiomic features has to be defined to assure the comparability between different studies (Yip and Aerts 2016). Variability of radiomic parameters was examined in PET (Galavis *et al* 2010, Tixier *et al* 2012, Leijenaar *et al* 2013, 2015, Yan *et al* 2015, Lu *et al* 2016, van Velden *et al* 2016), MRI (Mayerhoefer *et al* 2009, Maani *et al* 2013) and CT images (Cunliffe *et al* 2012, Duda *et al* 2013, Mackin *et al* 2015, Zhao *et al* 2016) in context of for example image reconstruction, discretization and registration. However, a similar evaluation for more computationally complex techniques, such as CTP, is still missing. The computation of perfusion maps from a repeated CT scan is performed over multiple steps and thus a radiomics stability analysis is of great importance for CTP (Fieselmann *et al* 2011). All available CTP software requires input from a user, such as contouring of the artery, voxel size or Hounsfield unit threshold adjustment. Vendors offer various approaches for automated artery contouring. However, it was shown that they introduce a 10%–65% failure rate due to motion artifacts and random noise (Kao *et al* 2014). The influence on the radiomics signature of the CTP computation factors, which cannot be standardized between the scans, is of great importance as they can introduce an error to any treatment response prognostic model.

Therefore, the aim of this study was to investigate the stability of radiomic parameters in the CTP maps regarding CTP calculation factors, CTP scanning protocol and image discretization to define new and stable parameters for treatment response prognostic models. Tumor volume, which is closely related to the UICC T-Stage, is a major prognostic factor for local tumor control after chemo-radiotherapy of lung cancer (Karstens *et al* 1990) and head and neck cancer (Studer and Glanzmann 2013). Different studies reported dependencies of radiomic parameters on tumor volume (Brooks and Grigsby 2014, Orlhac *et al* 2014, Hatt *et al* 2015). To show a potential added value of radiomic parameters in treatment response prognostic models over the known tumor volume related model, only the radiomic parameters uncorrelated with tumor volume were considered.

Materials and methods

Patient population and image acquisition

This retrospective analysis was performed on two cohorts of patients collected in prospective studies approved by a local ethical commission. Informed consent was obtained from all patients. Eleven patients with oropharyngeal carcinoma (OPC) and eleven patients with non-small cell lung cancer (NSCLC), who underwent tumor CT perfusion (CTP) imaging, were included in this analysis. The analyzed tumors were histologically confirmed stages T1–T4 and had median sizes of 21.8 cm^3 (9.5–42.3 cm^3) and 24.3 (5.8–224.2 cm^3) in OPC and NSCLC, respectively. The image acquisition was performed on two different CT scanners (Discovery VCT, GE Healthcare, Waukesha, WI for OPC and SOMATOM Definition Flash, Siemens Healthcare, Erlangen, Germany for NSCLC). The CTP scanning protocols suggested by the vendors were applied. Scanning parameters are presented in table 1. Iodinated contrast medium (40 ml of Ultravist® 370, Bayer Schering Healthcare, Germany) was injected at a flow of 7 ml s^{-1} for NSCLC and 5 ml s^{-1} in OPC (van Elmpt *et al* 2013, Veit-Haibach *et al* 2013). After a delay of 3–5 s, the perfusion scans were acquired. Images were reconstructed using filtered back projection. In the case of NSCLC the slice thickness was 3 mm and the in-plane resolution was in the range of 0.5×0.5–$1.0 \times 1.0\,\text{mm}^2$ dependent on patient size, whereas in the OPC scans the voxel size was fixed to $0.7 \times 0.7 \times 5.0\,\text{mm}^3$.

Table 1. CT perfusion image acquisition parameters for oropharyngeal carcinoma (OPC) and non-small cell lung cancer (NSCLC).

Scanning parameters	OPC	NSCLC
kV	100	80; 100
mAs	80	80; 100
Cranial-caudal coverage (cm)	4	7
Temporal resolution (s)	1	3
Number of images	45	60
Reconstruction increment (mm)	5	3
Convolution kernel	Standard	B30f
CT scanner	Discovery VCT, GE Healthcare, Waukesha, WI	SOMATOM Definition flash, Siemens Healthcare, Erlangen, Germany

Perfusion maps calculations

The pre-processing of CT images was performed before perfusion calculation. The in-plane resolution of the images was adjusted to the largest used scanning resolution of $1 \times 1\,\text{mm}^2$ by linear interpolation between pixels. To account for respiratory motion, lung scans were rigidly registered in tumor region using SmartAdapt® (v. 11, Varian Medical Systems, USA). If a tumor was blurred in one single scan, and thus no reliable registration was possible, the linear interpolation between two adjacent time points was used to replace the scan.

For the perfusion computation an in-house developed computer program based on the singular value decomposition method was used (Nesteruk *et al* 2015). Three perfusion maps were calculated: blood volume (BV), blood flow (BF) and mean transit time (MTT). These maps were normalized to the perfusion parameters inside the artery used for arterial input function (AIF) determination, the carotid artery and the aorta for OPC and NSCLC, respectively. A Hounsfield unit (HU) threshold was used to exclude bone, air and CT artifacts from the perfusion calculations (table 2).

Tumors were contoured by an experienced radiation oncologist (12 years of experience) in the anatomical CT scan, which was later used for perfusion calculation. The contours were copied on to perfusion maps. To exclude outliers a Gaussian function was fitted to the CTP intensities inside the tumor and points outside the 3-sigma limit were removed from further analysis. The radiomics signature was studied in these contoured tumor regions.

Radiomics analysis

The following 3D radiomic features extraction methods were implemented:

- statistical-based approach:
  - the histogram of intensities (HI) parameters ($n = 5$)
  - the gray level co-occurrence matrix (GLCM) parameters ($n = 14$) (Haralick *et al* 1973)
  - the neighborhood gray tone difference matrix (NGTDM) parameters ($n = 4$) (Amadasun and King 1989)
  - the gray level size zone matrix (GLSZM) parameters ($n = 11$) (Thibault *et al* 2009)
- model-based approach:
  - the fractal dimension ($n = 1$) (Napolitano *et al* 2012)
- transform-based approach

Table 2. Studied CT perfusion calculation and image discretization factors. The set of reference perfusion calculation factors used for separate evaluation of sources of radiomic parameters instability is shown in bold.

		Levels	
Type	CTP factors	OPC	NSCLC
Potentially standardizable factors	Image discretization (fixed number of bins)	16, 32, **64**	
	Image discretization (fixed intervals)	Blood volume and blood flow: 0.5%, 1%, 2% Mean transit time: 5%, 10%, 20%	
	Hounsfield unit (HU) intervals for exclusion of non-soft tissue from the analysis	lower threshold: **−20 HU**	lower threshold: −450 HU; −400 HU; −350 HU; **−300 HU**; −250 HU; −200 HU; −150 HU
		upper threshold: 120 HU; 140 HU; 160 HU; **180 HU**; 200 HU; 220 HU; 240 HU	upper threshold: **200 HU**
	Voxel size (mm^3)	1 × 1 × 5; 2 × 2 × 5; 3 × 3 × 5; 4 × 4 × 5; **5 × 5 × 5**	1 × 1 × 3; 2 × 2 × 3; **3 × 3 × 3**; 4 × 4 × 3; 5 × 5 × 3
	Temporal resolution (s)	**1**, 2, 3	**3**, 6
Non-standardizable factors	Artery contouring (AIF)	Perfusion maps calculated based on 5 different contours of the artery	
	Noise threshold in perfusion maps calculation	10%, 12%, 14%, 16%, 18%, **20%**, 22%, 24%, 26%, 28%, 30%	

Table 3. The list of stable and uncorrelated radiomic parameters after standardization of image discretization, HU threshold, voxel size and temporal resolution.

Perfusion map	Stable parameter
Blood flow	entropy, sum entropy, LHH low gray-level zone emphasis
Mean transit time	large zone low gray-level emphasis
Blood volume	difference entropy, coarseness, large zone high gray-level emphasis, HLH information measure of correlation 2, LLL coefficient of variation, LLL average

The full list of the radiomic features is presented in the supplement (supplementary table 3 (stacks.iop.org/PMB/61/8736/mmedia)). The wavelet transform was applied to enhance details and edges in the maps. The 35 radiomic parameters (from statistical- and model-based approach) were computed for a perfusion map and its 3D wavelet transform (8 sub-bands). It summed up to 315 radiomic features per each perfusion map (BV, BF and MTT) and together 945 radiomic features per patient. Radiomics analysis was implemented in Python programming language (v. 2.7). The wavelet decomposition was performed using PyWavelets library and the 'Coif1' wavelet as suggested in Aerts *et al* (2014). The details of the calculations are presented in the supplement (section 2).

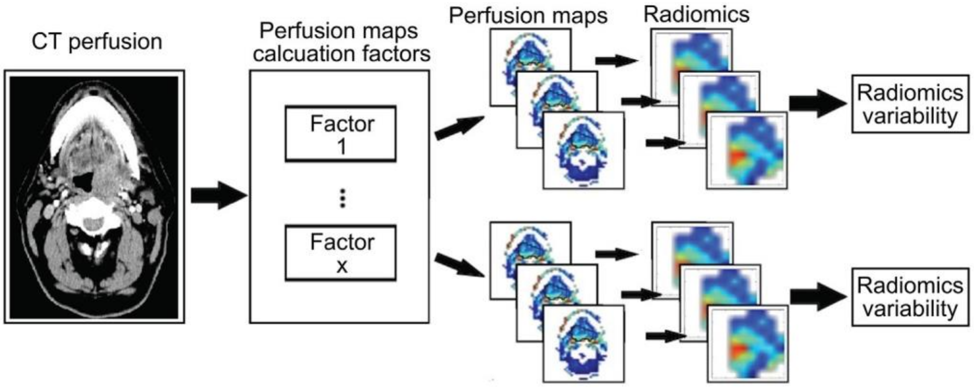


Figure 1. Scheme of the radiomics variability analysis. Perfusion maps were calculated using multiple values for 7 perfusion calculation factors. Radiomics variability was studied separately for each of the factors.

Stability of radiomic parameters in CTP

The stability of radiomic parameters was studied in regard to perfusion maps calculation factors and image discretization. The scheme of the stability analysis is presented in figure 1. The list of investigated factors is presented in table 2. They were divided into factors possibly standardizable among different scans (image discretization, HU threshold, voxel size and CT temporal resolution) and non-standardizable factors (artery contouring and noise level). A set of the following reference perfusion computation factors was defined. For higher-order statistical analysis and transformed-based analysis, images were discretized into 64 bins, which assured the visualization of the inhomogeneity in the image while keeping the noise at the low level. The reference HU range allowed the best visual separation of the tumor and overlapping or adjacent non-perfused tissues/materials. Definitions of radiomic parameters were developed assuming square pixels, therefore in this analysis the cubic voxels were used as the reference. Their size was adapted to the smallest resolution in the image, which corresponded to the slice thickness. Original temporal scanning resolution was chosen as the reference. A recommended noise threshold of 20% in perfusion calculations using singular value decomposition was included in the reference set (Fieselmann *et al* 2011). As the reference AIF definition the artery contour with the least noisy enhancement curve was chosen.

To investigate the stability of the radiomic features in the perfusion maps in respect to a particular factor, the maps were calculated multiple times using this reference set and changing the factor of interest. The analysis was comprised of four main parts:

(1) The variability of radiomic parameters with respect to the image discretization method (fixed number of bins in comparison to fixed intervals) was studied using the consistency in the intraclass correlation (ICC) two-way mixed model (IBM SPSS Statistics for Windows, v. 22.0, USA).
(2) The Spearman correlations of tumor volume with radiomic parameters were investigated in the perfusion maps calculated using the set of reference perfusion computation factors.

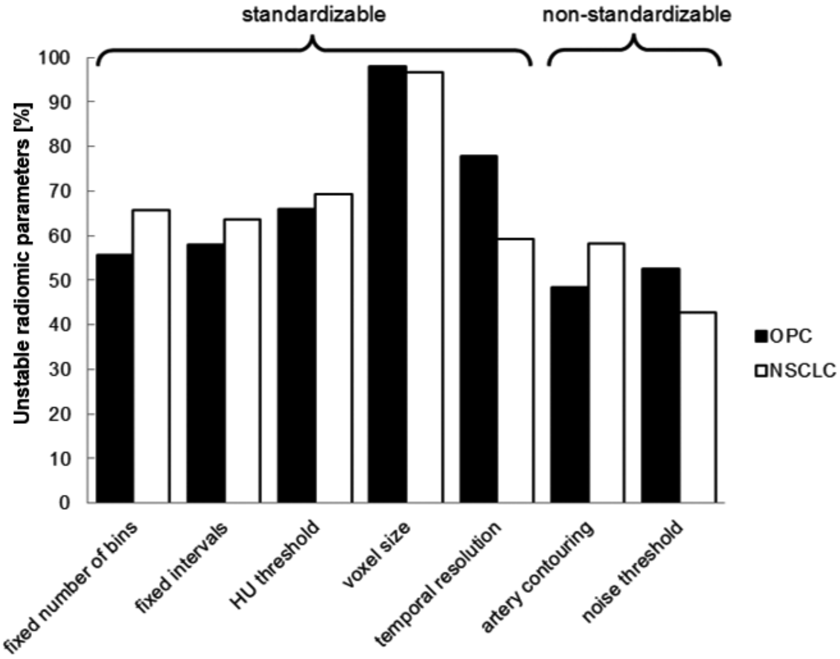


Figure 2. Percentage of unstable radiomic parameters regarding different perfusion calculation factors and image discretization. OPC—oropharyngeal cancer, NSCLC—non-small cell lung cancer.

(3) To further examine the stability of radiomic parameters the ICC was calculated for all other factors influencing the perfusion maps computation, except image discretization (table 2). The stability was first analyzed according to tumor entity (OPC or NSCLC) and only the features stable in both entities were included in the final set.
(4) The stable parameters were grouped according to inter-parameter Spearman correlations and only the parameter with the highest ICC was chosen.

The acceptance level was 0.9 and 0.7 for the ICC and Spearman correlation, respectively. Despite the small cohorts of patients, the high acceptance level in ICC allowed to keep the type I and type II errors at the 0.05 and 0.20 levels, respectively (Walter *et al* 1998).

Results

Stability of radiomic parameters in respect to image discretization method

The two discretization techniques, with either fixed number of bins or fixed intervals, were evaluated using the ICC. They resulted in a similar number of stable radiomic parameters (figure 2). The radiomic parameters calculated in the MTT maps and their wavelet transforms were the least affected by image discretization. For further analysis, a fixed number of bins method was chosen.

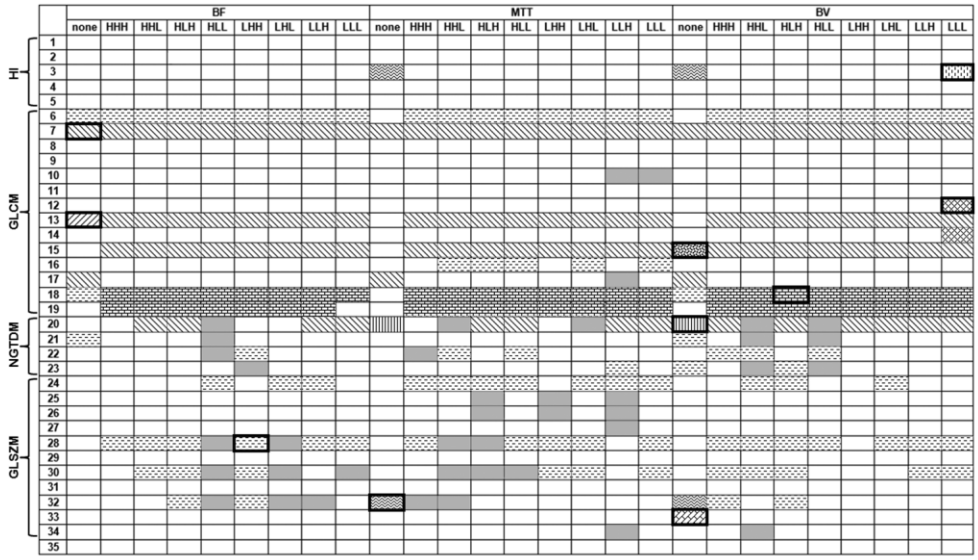


Figure 3. Relationships between the studied radiomic parameters. Parameters, which were excluded because of correlation with tumor size, are marked in gray. Inter-parameter correlation groups are marked with the same texture; the most stable parameter in each group is additionally marked with a black frame. The white rectangles correspond to parameters, which were not stable regarding potentially standardizable factors. Columns correspond to different types of images: BF—blood flow, MTT—mean transit time, BV—blood volume, none—original image, HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL—3D wavelet transforms. Rows correspond to different radiomic parameters: HI—histogram of intensities, GLCM—gray level co-occurrence matrix, NGTDM—neighborhood gray tome difference matrix, GLSZM—gray level size zone matrix, 1—mean, 2—standard deviation, 3—coefficient of variation, 4—skewness, 5—kurtosis, 6—energy, 7—entropy, 8—contrast, 9—correlation, 10—homogeneity, 11—variance, 12—average, 13—sum of entropy, 14—sum of variance, 15—difference entropy, 16—difference variance, 17—information measure of correlation 1, 18—information measure of correlation 2, 19—maximal correlation coefficient, 20—coarseness, 21—contrast, 22—busyness, 23—complexity, 24—gray-level non-uniformity, 25— size zone non-uniformity, 26—small zone emphasis, 27—large zone emphasis, 28— low gray-level zone emphasis, 29—high gray-level zone emphasis, 30—small zone low gray-level emphasis, 31—small zone high gray-level emphasis, 32—large zone low gray-level emphasis, 33—large zone high gray-level emphasis, 34—zone percentage 35—fractal dimension.

Correlation of radiomic parameters and tumor volume

Forty radiomic parameters correlated strongly with tumor volume (Spearman correlation $r > 0.7$, $p < 0.05$) and were therefore excluded from further analysis. Most of these parameters were calculated based on the NGTDM and the GLSZM (figure 3). None of the 105 radiomic parameters calculated in the 3 non-transformed maps was correlated with tumor volume.

Stability of radiomic parameters

The stability of radiomic parameters was investigated in respect to perfusion maps calculation factors and image discretization (fixed number of bins) using ICC. The scheme showing the

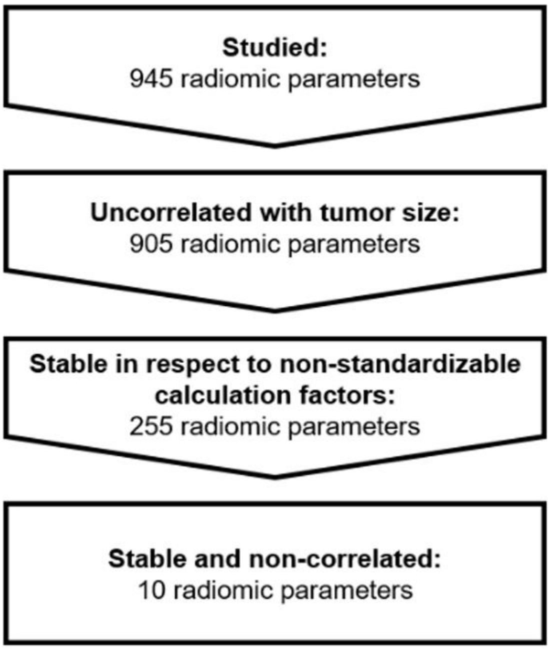


Figure 4. Scheme of exclusion steps performed for the stability analysis. The numbers of stable radiomic parameters resulting from analysis of the mixed group of OPC and NSCLC patients are shown.

steps of analysis and the corresponding number of stable radiomic parameters is presented in figure 4. The potentially standardizable factors introduced more variability into radiomic features than the non-standardizable ones (figure 2). The highest variability was observed for voxel size. It caused instability of 98% and 97% of the radiomic parameters in OPC and NSCLC, respectively.

A similar percentage of radiomic parameters was unstable in the OPC and NSCLC in respect to a particular perfusion computation factor. The highest discrepancy between OPC and NSCLC was observed for temporal resolution, which could be caused by a wider range of studied temporal resolution in the OPC in comparison to NSCLC.

Information measure of correlation 1 in MTT maps was the only stable radiomic parameter in OPC against all studied factors (standardizable and non-standardizable). Nine other radiomic parameters were stable in NSCLC (supplementary table 1). None of the radiomic parameters was found stable for both tumor entities. To define radiomic parameters that are stable regardless of tumor entity, the final set was defined assuming the standardization of image discretization, HU threshold, voxel size and temporal resolution according to the reference set.

Inter-parameter correlations

After standardization of potentially standardizable perfusion calculation factors 382, 338 and 255 radiomic parameters were stable in OPC, NSCLC and both tumor entities, respectively. The average ICC for each stable radiomic parameter is shown in the supplement (supplementary figure 1). The number of groups found after accounting for the inter-parameter Spearman correlation is shown in supplementary table 2. More groups were found for OPC in comparison to NSCLC. Only 10 separate groups were identified in both tumor entities, OPC and NSCLC (figure 3). The size of these groups varied from 1 to 91 parameters. From each group the parameter with the highest ICC was chosen. This resulted in a set of uncorrelated radiomic parameters, which were stable in respect to non-standardizable perfusion calculation factors, and were not correlated with tumor volume (table 3). The final groups of correlated radiomic parameters depend on the parameters that are stable in the studied cohorts (OPC, NSCLC and both tumor entities). Therefore, the final parameters stable in both tumor entities not always correspond to the final parameters stable per tumor entity.

Discussion

This study evaluated the stability of radiomic parameters in CT perfusion maps for two tumor entities, oropharyngeal carcinoma (OPC) and non-small cell lung cancer (NSCLC), in respect to perfusion calculation factors and image discretization. The non-standardizable factors (artery contouring and noise level) influenced the stability of fewer radiomic parameters than the potentially standardizable ones (ranges of unstable parameters 56%–98% and 43–58%, respectively). The lack of any standardization would result in instability of all radiomic parameters, for both tumor entities. Thus they could not be used as an input in a treatment response prognostic model. However, a set of 10 uncorrelated and stable radiomic parameters was identified after the standardization of image discretization, HU threshold, voxel size and temporal resolution. Each of the methods used in the statistical-based radiomics analysis (the histogram of intensities, the GLCM, the NGTDM and the GLSZM) was represented in the final feature set. Radiomic features calculated in all three perfusion maps can be found in the final set. The BV map was characterized by the biggest number of stable radiomic parameters.

The aim of this study was to identify new imaging parameters for future prognostic models. Tumor volume that is closely related to the UICC T-stage is a major prognostic factor for local tumor control after chemo-radiotherapy (Karstens *et al* 1990, Studer and Glanzmann 2013). It was reported in the literature that some of the radiomic parameters are correlated with tumor volume (Brooks and Grigsby 2014, Orlhac *et al* 2014, Hatt *et al* 2015). In this study 40 parameters closely correlated with tumor volume were excluded from further stability analysis with the aim to create a model with an added value over the classical volume-related one.

More radiomic parameters were found to be stable in the OPC in comparison to NSCLC regarding non-standardizable factors. Multiple contouring of the supplying artery resulted in instability of 567 parameters in NSCLC and only 480 in OPC (supplementary figure 2). The lower resolution of OPC maps and small diameter of carotid artery could have caused smaller variation in the artery input function and as a result a smaller variation in tumor perfusion. On the contrary, OPC maps were more sensitive to changes in the noise threshold, which could be caused by a smaller number of voxels used for the radiomic parameters calculation. Due to a smaller tumor size and a worse image resolution the median number of voxels used in the analysis was 900 in NSCLC and only 174 in OPC. The sensitivity of OPC to noise changes caused for example instability of the mean perfusion parameter (supplementary figure 3). Some differences in stability of radiomic parameters between tumor sites could be also caused by the use of two CT scanners from two manufactures. Mackin *et al* 2015 showed a substantial influence of scanner type on the radiomics stability in CT. However, this study was performed on a limited number of radiomic parameters and using 2.5D analysis (Zhang *et al* 2015). Instead of an entity-specific or scanner-specific analysis resulting in a bigger number of stable radiomic parameters, the combined analysis of the data from two tumor entities was performed to increase the applicability of the stable radiomic parameters as future prognostic factors.

The radiomic features stability was investigated in respect to the defined reference set of the calculation factors. Image discretization into 64 bins was defined as a reference instead of a fixed-interval approach despite the results from recent studies in [18F]-FDG PET images (Leijenaar *et al* 2015, van Velden *et al* 2016). It was shown that the fixed-interval technique, based on the standardized uptake value, resulted in smaller variations in the radiomics signature in comparison to the fixed number of bins also in terms of delineation technique and image reconstruction. However, these results refer only to non-transformed images. In addition, the definitions of radiomic parameters are not normalized to the number of bins used in the analysis. In our CTP study both discretization techniques performed equally good in terms of radiomic parameter stability. Considering the arguments listed above, we decided that the fixed number of bin discretization is more suitable for our analysis.

The choice of voxel size is critical for radiomic features determination. A large variation of radiomic features in respect to voxel size was reported in (18F)-FDG PET images (Galavis *et al* 2010). Among all studied image reconstruction parameters (reconstruction algorithm, iteration numbers, grid size and post-filtering) the grid size was found to cause the highest instability of radiomic features (Galavis *et al* 2010). Similar results were observed in this work on CTP. Definitions of radiomic parameters assume the symmetry of voxels and thus, in this study, the set of reference factors included the symmetrical voxels despite their low resolution.

Single CTP examination in head and neck cancer delivers around 0.2 Gy to the patient, thus dose reduction is an important topic in CTP (Ramirez-Grialdo 2013, Niesten *et al* 2014). The increase of temporal resolution of the scans is a potential solution. However, it was proven to cause large changes in the image intensities, underestimation of MTT and overestimation of BV and BF (Wintermark *et al* 2004). In the radiomics analysis of CTP the instability of intensity histogram-based parameters was observed in respect to different temporal resolutions. Nevertheless, some of the higher order radiomic parameters were found to be stable. A treatment response prognostic model based on higher order radiomic parameters could thus allow both for an increase in temporal resolution and a dose reduction.

Among 945 studied radiomic parameters 255 were found to be stable in both tumor entities. However, most of them were correlated with each other and thus created groups. In some of the groups the differences in the ICC between different radiomic parameters were minor. Despite the fact that only the parameters with highest ICC were included in the final set, all the parameters in the groups are good candidates for the input to a prognostic model. Further studies are needed to identify the strongest prognostic parameters in these groups.

Several studies already showed the instability of the radiomics signature in medical images in respect to acquisition and post-processing protocols (Mayerhoefer *et al* 2009, Galavis *et al* 2010, Cunliffe *et al* 2012, Duda *et al* 2013, Leijenaar *et al* 2013, 2015, Maani *et al* 2013, Yan *et al* 2015, van Velden *et al* 2016). The need for standardization of image discretization and voxel size was shown to be of a great importance in PET images (Galavis *et al* 2010, Leijenaar *et al* 2015, Yan *et al* 2015, Lu *et al* 2016). This study showed that in CTP maps additionally to the standardization of these two factors the standardization of HU intervals and temporal resolution is recommended.

There are some additional factors, which may influence tumor radiomics signature in CTP, but were not evaluated in this study. E.g. it is well known that perfusion estimation depends on the used calculation model (Kudo *et al* 2010) and its influence on radiomic parameters needs to be further studied. Additionally, inter-observer variability in tumor contouring was shown to influence tumor radiomic features stability in PET and CT images (Leijenaar *et al* 2013, Parmar *et al* 2014, Lu *et al* 2016). In this study tumors were contoured by an experienced radiation oncologist and inter-observer variability was not considered.

Conclusion

CT perfusion calculation factors greatly influence the stability of radiomic parameters. Voxel size, image discretization, HU intervals and temporal resolution have to be standardized to build a reliable prognostic model based on CTP radiomics analysis. A set of 10 stable radiomic parameters in CTP was identified for two tumor entities. This marker set will next be evaluated in clinical correlative studies for its propensity to predict treatment outcome.

Acknowledgments

The project was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant of the Center for Clinical Research, University and University Hospital Zurich and by a research grant from Merck (Schweiz) AG. The authors declared no potential conflict of interest with respect to the research, authorship, and/or publication of this paper.

References

Aerts H J W L *et al* 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach *Nat. Commun.* **5** 4006

Amadasun M and King R 1989 Texural features corresponding to textural properties *IEEE Trans. Syst. Man Cybern.* **19** 1264–74

Brooks F J and Grigsby P W 2014 The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake *J. Nucl. Med.* **55** 37–42

Cunliffe A R, Al-Hallaq H A, Labby Z E, Pelizzari C A, Straus C, Sensakovic W F, Ludwig M and Armato S G III 2012 Lung texture in serial thoracic CT scans: assessment of change introduced by image registration *Med. Phys.* **39** 4679–90

Duda D, Kretowski M and Bezy-Wendling J 2013 Effect of slice thickness on texture-based classification of liver dynamic CT scans *Computer Information Systems and Industrial Management* (Berlin: Springer) pp 96–107

Fieselmann A, Kowarschik M, Ganguly A, Hornegger J and Fahrig R 2011 Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details *Int. J. Biomed. Imaging* **2011** 467563

Galavis P A, Hollensen C, Jallow N, Paliwal B and Jeraj R 2010 Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters *Acta Oncol.* **49** 1012–6

Goh V, Halligan S, Daley F, Wellsted D M, Guenther T and Bartram C I 2008 Colorectal tumor vascularity: quantitative assessment with multidetector CT—do tumor perfusion measurements reflect angiogenesis? *Radiology* **249** 510–7

Haralick R M, Shanmugam K and Dinstein I 1973 Textural features for image classification *IEEE Trans. Syst. Man Cybern.* **3** 610–21

Hatt M *et al* 2015 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort *J. Nucl. Med.* **56** 38–44

Itakura H *et al* 2015 Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities *Sci. Transl. Med.* **7** 303ra138

Kao Y H, Mu Huo Teng M, Kao Y T, Chen Y J, Wu C H, Chen W C, Chiu F Y and Chang F C 2014 Automatic measurements of arterial input and venous output functions on cerebral computed tomography perfusion images: a preliminary study *Comput. Biol. Med.* **51** 51–60

Karstens J H, Andreopoulos D and Ammomn J 1990 Initial tumor size and local control in stage III non-small cell lung cancer terated by radio-chemotherpay *Onkologie* **13** 144–5

Kudo K, Sasaki M, Yamada K, Momoshima S, Utsunomiya H, Shirato H and Ogasawara K 2010 Differences in CT perfusion maps generated by different commercial software: quantitative analysis by using identical source data of acute stroke patients *Radiology* **254** 200–9

Kumar M and Mehta K K 2011 A texture based tumor detection and automatic segmentation using seeded region growing method *Int. J. Comput. Technol. Appl.* **2** 855–9

Lamlertthon W, Hayward M C and Hayes D N 2011 Emerging technologies for improved stratification of cancer patients: a review of opportunities, challenges, and tools *Cancer J.* **17** 451–64

Lee T Y, Purdie T G and Stewart E 2003 CT imaging of angiogenesis *Q. J. Nucl. Med.* **47** 171–87

Leijenaar R T H *et al* 2013 Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability *Acta Oncol.* **52** 1391–7

Leijenaar R T H, Nalbantov G, Carvalho S, van Elmpt W J C, Troost E G C, Boellaard R, Aerts H J W L, Gillies R J and Lambin P 2015 The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis *Sci. Rep.* **5** 11075

Lin L, Yang W, Li C, Tang J and Cao X 2016 Inference with collaborative model for interactive tumor segmentation in medical image sequences *IEEE Trans. Cybern.* **46** 2796–809

Ling S, Deng D, Mo Y, Zhang X, Guan X and Wei Q 2014 Correlations between CT perfusion parameters and vascular endothelial growth factor expression and microvessel density in implanted VX2 lung tumors *Cell Biochem. Biophys.* **70** 629–33

Lu L, Lv W, Jiang J, Ma J, Feng Q, Rahmim A and Chen W 2016 Robustness of radiomic features in [11C]Choline and [18F]FDG PET/CT imaging of nasopharyngeal carcinoma: impact of segmentation and discretization *Mol. Imaging Biol.* **18** 935–45

Luczynska E, Gasinska A, Blecharz P, Stelmach A, Jereczek-Fossa B A and Reinfuss M 2014 Value of perfusion CT parameters, microvessl density and VEGF expression in differentiation of benign and malignant prostate tumours *Pol. J. Pathol.* **65** 229–36

Maani R, Kalra S and Yang Y H 2013 Noise robust rotation invariant features for texture classification *Pattern Recognit.* **46** 2103–16

Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, Rodriguez-Rivera E, Dodge C, Jones A K and Court L 2015 Measuring computed tomography scanner variability of radiomics features *Investigative Radiol.* **50** 757–65

Materka A 2004 Texture analysis methodologies for magnetic resonance imaging *Dialogues Clin. Neurosci.* **6** 243–50

Mayerhoefer M E, Szomolanyi P, Jirak D, Materka A and Trattnig S 2009 Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study *Med. Phys.* **36** 1236–43

Napolitano A, Ungania S and Cannata V 2012 Fractal dimension estimation methods for biomedical images *MATLAB—a Fundamental Tool for Scientific Computing and Engineering Applications* vol 3 (Rijeka: InTech)

Nesteruk M *et al* 2015 Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma *Radiother. Oncol.* **117** 125–31

Niesten J M, van der Schaaf I C, Riordan A J, de Jong H W, Horsch A D, Eijspaart D, Smit E J, Mali W P and Velthuis B K 2014 Radiation dose reduction in cerebral CT perfusion imaging using iterative reconstruction *Eur. Radiol.* **24** 484–93

Orlhac F, Soussan M, Maisonobe J A, Garcia C A, Vanderlinden B and Buvat I 2014 Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis *J. Nucl. Med.* **55** 414–22

Parmar C *et al* 2014 Robust Radiomics feature quantification using semiautomatic volumetric segmentation *PLoS One* **9** e102107

Parmar C, Leijenaar R T H, Grossmann P, Velazquez E R, Bussink J, Rietveld D, Rietbergen M M, Haibe-Kains B, Lambin P and Aerts H J W L 2015 Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer *Sci. Rep.* **5** 11044

Preda L, Calloni S F, Moscatelli M E, Cossu Rocca M and Bellomi M 2014 Role of CT perfusion in monitoring and prediction of response to therapy of head and neck squamous cell carcinoma *Biomed. Res. Int.* **2014** 917150

Ramirez-Grialdo J C 2013 Evaluation of strategies to reduce radiation dose in perfusion CT imaging using reproducilbe biologic phantom *AJR Am. J. Roentgenol.* **200** 621–7

Rao A R and Lohse G L 1993 Identifying high level features of texture perception *Graph. Model. Im. Proc.* **55** 218–33

Studer G and Glanzmann C 2013 Volumetric stratification of cT4 stage head and neck cancer *Strahlenther. Onkol.* **189** 867–73

Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J and Mari J L 2009 Texture indexes and gray level size zone matrix application to cell nuclei classification *Pattern Recognit. Inf. Process* **2009** 140–5

Tixier F, Hatt M, Le Rest C C, Le Pogam A, Corcos L and Visvikis D 2012 Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET *J. Nucl. Med.* **53** 693–700

van Elmpt W *et al* 2013 Characterization of tumor heterogeneity using dynamic contrast enhanced CT and FDG-PET in non-small cell lung cancer *Radiother. Oncol.* **109** 65–70

van Velden F H, Kramer G M, Frings V, Nissen I A, Mulder E R, de Langen A J, Hoekstra O S, Smit E F and Boellaard R 2016 Repeatability of radiomic features in non-small-cell lung cancer [(18)F] FDG-PET/CT studies: impact of reconstruction and delineation *Mol. Imaging Biol.* **18** 788–95

Vaupel P and Mayer A 2007 Hypoxia in cancer: significance and impact on clinical outcome *Cancer Metastasis Rev.* **26** 225–39

Veit-Haibach P, Schmid D, Strobel K, Soyka J D, Schaefer N G, Haerle S K, Huber G, Studer G, Seifert B and Hany T F 2013 Combined PET/CT-perfusion in patients with head and neck cancers *Eur. Radiol.* **23** 163–73

Walter S D, Eliasziw M and Donner A 1998 Sample size and optimal designs for reliability studies *Stat. Med.* **17** 101–10

Wintermark M, Smith W S, Ko N U, Quist M, Schnyder P and Dillon W P 2004 Dynamic perfusion CT: optimizing the temporal resolution and contrast volume for calculation of perfusion CT parameters in stroke patients *AJNR Am. J. Neuroradiol.* **25** 720–9

Yan J, Chu-Shern J L, Loi H Y, Khor L K, Sinha A K, Quek S T, Tham I W and Townsend D 2015 Impact of image reconstruction settings on texture features in 18F-FDG PET *J. Nucl. Med.* **56** 1667–73

Yang D, Rao G, Martinez J, Veeraraghavan A and Rao A 2015 Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12 months survival status in glioblastoma *Med. Phys.* **42** 6725–35

Yeung T P, Wang Y, He W, Urbini B, Gafà R, Ulazzi L, Yartsev S, Bauman G, Lee T Y and Fainardi E 2015 Survival prediction in high-grade gliomas using CT perfusion imaging *J. Neurooncol.* **123** 93–102

Yip S S F and Aerts H J W L 2016 Applications and limitations of radiomics *Phys. Med. Biol.* **61** 150–66

Zhang L, Fried D V, Fave X J, Hunter L A, Yang J and Court L E 2015 IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics *Med. Phys.* **42** 1341–53

Zhao B, Tan Y, Tsai W Y, Qi J, Xie C, Lu L and Schwartz L H 2016 Reproducibility of radiomics for deciphering tumor phenotype with imaging *Sci. Rep.* **6** 23428

Supplementary Information
Stability of radiomics features in CT perfusion maps

Marta Nesteruk[1*], Oliver Riesterer[1], Ralph A Bundschuh[2], Patrick Veit-Haibach[3,4], Martin Hüllner[3], Gabriela Studer[1], Sonja Stieb[1], Stefan Glatz[1], Martin Pruschy[1], Matthias Guckenberger[1], and Stephanie Tanadini–Lang[1]

[1] University Hospital Zurich and University of Zurich, Department of Radiation Oncology, Zurich, 8091, Switzerland
[2] University Hospital Bonn, Department of Nuclear Medicine, Bonn, 53127, Germany
[3] University Hospital Zurich and University of Zurich, Department of Nuclear Medicine, Zurich, 8091, Switzerland
[4] University Hospital Zurich and University of Zurich, Department of Diagnostic and Interventional Radiology, Zurich, 8091, Switzerland
[*] marta.nesteruk@usz.ch

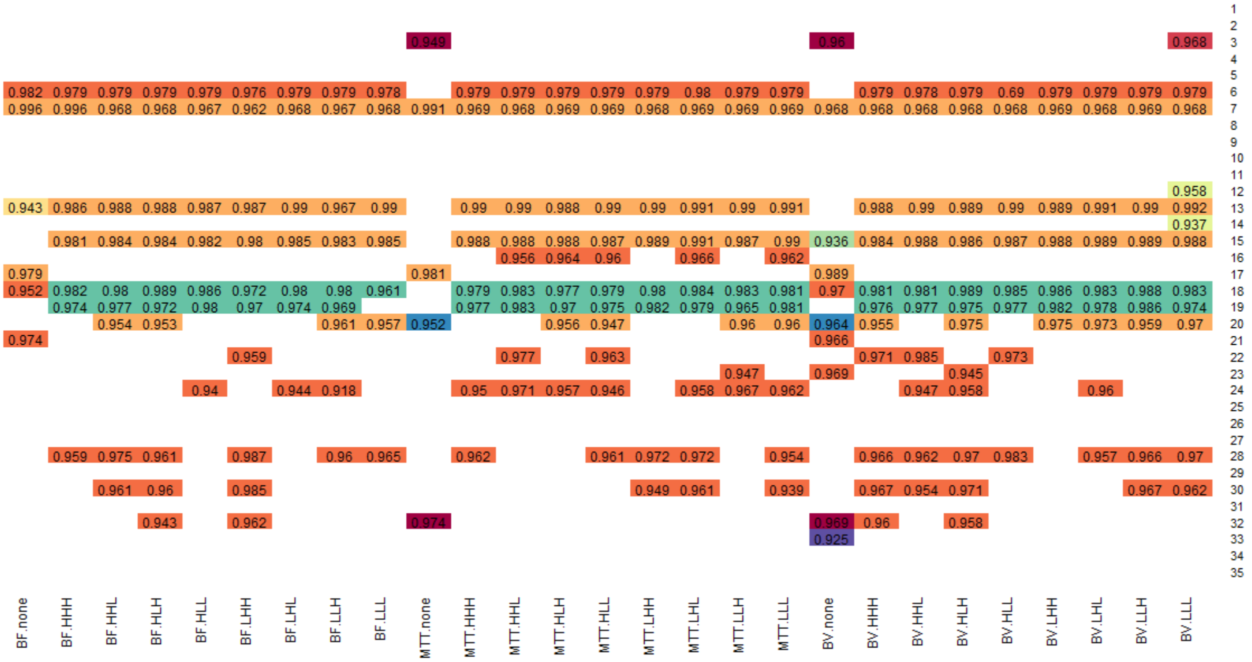
Results

Tumor entity	Perfusion map	Stable parameter
OPC	Blood flow	
	Mean transit time	Information measure of correlation 1
	Blood volume	
NSCLC	Blood flow	LLL mean
	Mean transit time	HLH Information measure of correlation 1, LHH Information measure of correlation 1, LLL entropy
	Blood volume	mean, LHL Information measure of correlation 1, LLL mean, LLL covariance, LLL Information measure of correlation 1

Table S1. List of the stable texture parameters. In the case of NSCLC stable parameters created four groups of correlated parameters. In each group the most stable parameter was underlined.

Tumor entity	Number of groups	Perfusion map	Stable parameter
OPC	34	Blood flow	covariance, sum entropy, HHH kurtosis, HHH average, HLH average, LHH skewness, LLL average
		Mean transit time	fractal dimension, HLH average, HLL mean, LHH maximal correlation coefficient, LHL average, LLL busyness, LLL fractal dimension
		Blood volume	covariance, kurtosis, energy, entropy, difference entropy, coarseness, low gray-level size emphasis, long size high gray-level emphasis, HHL skewness, HHL average, HLH average, HLL mean, LHH mean, LHH average, LHL mean, LHL average, LHH average, LLL standard deviation, LLL covariance, LLL average
NSCLC	13	Blood flow	mean, variance, sum entropy, HLH information measure of correlation 2
		Mean transit time	LLL entropy
		Blood volume	mean, correlation, difference entropy, information measure of correlation 2, coarseness, long size high gray-level emphasis, HLL long ize high gray-level emphasis, LHL long size emphasis

Table S2. List of the stable and uncorrelated texture parameters after standarization of image discretization, HU threshold, voxel size and temporal resolution. OPC – oropharyngeal carcinoma, NSCLC – non–small cell lung cancer.



BF.none	BF.HHH	BF.HHL	BF.HLH	BF.HLL	BF.LHH	BF.LHL	BF.LLH	BF.LLL	MTT.none	MTT.HHH	MTT.HHL	MTT.HLH	MTT.HLL	MTT.LHH	MTT.LHL	MTT.LLH	MTT.LLL	BV.none	BV.HHH	BV.HHL	BV.HLH	BV.HLL	BV.LHH	BV.LHL	BV.LLH	BV.LLL	
																											1
																											2
									0.949									0.96								0.968	3
																											4
																											5
0.982	0.979	0.979	0.979	0.979	0.976	0.979	0.979	0.978		0.979	0.979	0.979	0.979	0.979	0.98	0.979	0.979		0.979	0.978	0.979	0.69	0.979	0.979	0.979	0.979	6
0.996	0.996	0.968	0.968	0.967	0.962	0.968	0.967	0.968	0.991	0.969	0.968	0.969	0.969	0.968	0.969	0.969	0.969	0.968	0.968	0.968	0.968	0.968	0.969	0.968	0.969	0.968	7
																											8
																											9
																											10
																											11
																										0.958	12
0.943	0.986	0.988	0.988	0.987	0.987	0.99	0.967	0.99		0.99	0.99	0.988	0.99	0.99	0.991	0.99	0.991		0.988	0.99	0.989	0.99	0.989	0.991	0.99	0.992	13
																										0.937	14
	0.981	0.984	0.984	0.982	0.98	0.985	0.983	0.985		0.988	0.988	0.988	0.987	0.989	0.991	0.987	0.99	0.936	0.984	0.988	0.986	0.987	0.988	0.989	0.989	0.988	15
											0.956	0.964	0.96		0.966		0.962										16
0.979									0.981									0.989									17
0.952	0.982	0.98	0.989	0.986	0.972	0.98	0.98	0.961		0.979	0.983	0.977	0.979	0.98	0.984	0.983	0.981	0.97	0.981	0.981	0.989	0.985	0.986	0.983	0.988	0.983	18
	0.974	0.977	0.972	0.98	0.97	0.974	0.969			0.977	0.983	0.97	0.975	0.982	0.979	0.965	0.981		0.976	0.977	0.975	0.977	0.982	0.978	0.986	0.974	19
		0.954	0.953				0.961	0.957	0.952			0.956	0.947			0.96	0.96	0.964	0.955		0.975		0.975	0.973	0.959	0.97	20
0.974																		0.966									21
					0.959						0.977		0.963						0.971	0.985		0.973					22
																0.947		0.969			0.945						23
				0.94		0.944	0.918			0.95	0.971	0.957	0.946		0.958	0.967	0.962			0.947	0.958			0.96			24
																											25
																											26
																											27
	0.959	0.975	0.961		0.987		0.96	0.965		0.962			0.961	0.972	0.972		0.954		0.966	0.962	0.97	0.983		0.957	0.966	0.97	28
																											29
		0.961	0.96		0.985									0.949	0.961		0.939		0.967	0.954	0.971				0.967	0.962	30
																											31
			0.943		0.962				0.974									0.969	0.96		0.958						32
																		0.925									33
																											34
																											35

Figure S1. The average intraclass correlation (ICC) in respect to artery contouring and noise threshold (potentially non–standardizable factors). The correlated texture parameters are marked with the same color. The ICC is given only for the stable parameters.

Columns correspond to different types of images: BF – blood flow, MTT – mean transit time, BV – blood volume, none – original image, HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL – 3D wavelet transforms, H – high–pass filter, L – low–pass filter.

Rows correspond to different texture parameters: HI – histogram of intensities, GLCM – Gray Level Co-occurrence Matrix, NGTDM – Neighborhood Gray Tome Difference Matrix, GLSZM – Gray Level Size Zone Matrix, 1 – mean, 2 – standard deviation, 3 – covariance, 4 – skewness, 5 – kurtosis, 6 – energy, 7 – entropy, 8 – contrast, 9 – correlation, 10 – homogeneity, 11 – variance, 12 – average, 13 – sum of entropy, 14 – sum of variance, 15 – difference entropy, 16 – difference variance, 17 – information measure of correlation 1, 18 – information measure of correlation 2, 19 – maximal correlation coefficient, 20 – coarseness, 21 – contrast, 22 – busyness, 23 – complexity, 24 – gray–level non–uniformity, 25 – size zone non–uniformity, 26 – short size emphasis, 27 – long size emphasis, 28 – low gray–level size emphasis, 29 – high gray–level size emphasis, 30 – short size low gray–level emphasis, 31 – short size high gray–level emphasis, 32 – long size low gray–level emphasis, 33 – long size high gray–level emphasis, 34 – size percentage 35 – fractal dimension.

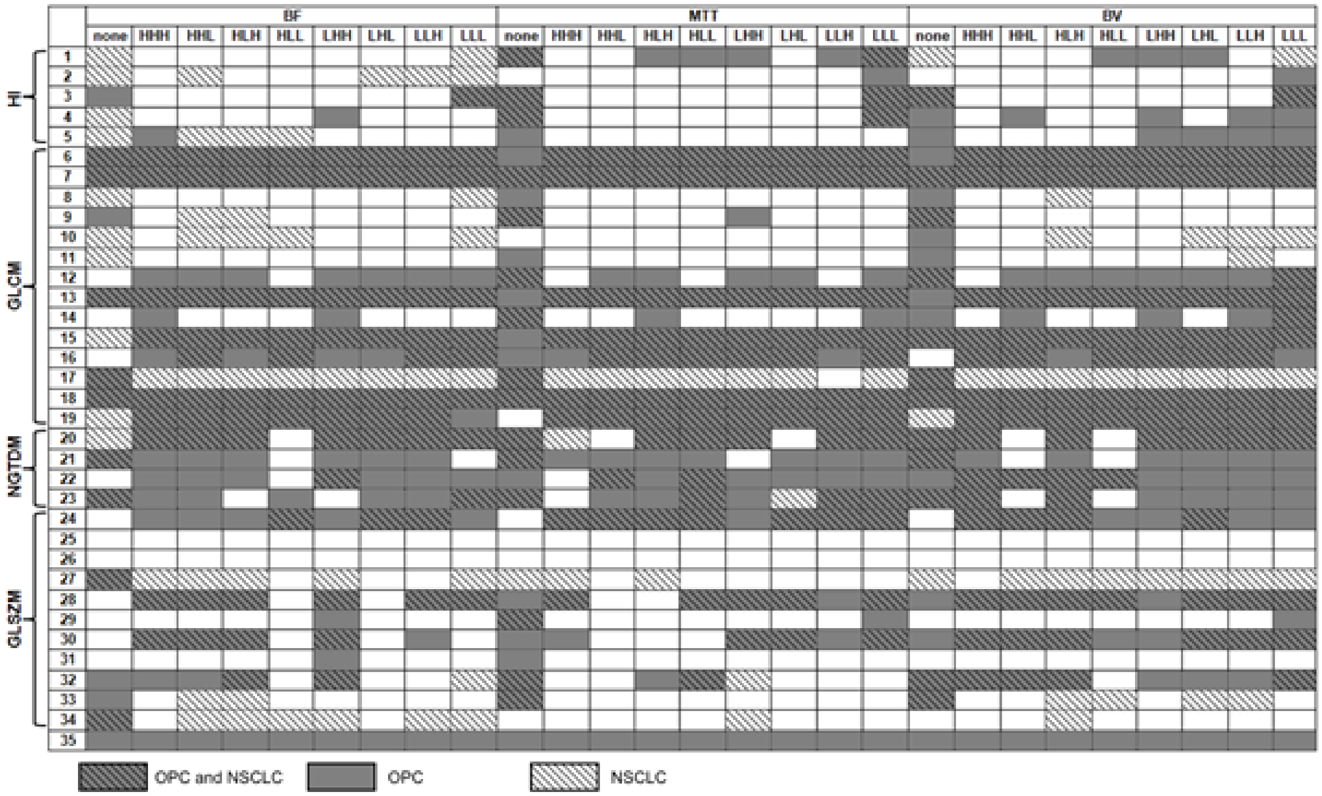


Figure S2. Robustness of texture parameters regarding artery contouring . The mean values of blood volume and blood flow (row number 1) were stable for non–small cell lung cancer (NSCLC) but not for oropharyngeal cancer (OPC) in respect to artery contouring. Changes in the noise caused instability of the mean MTT in both tumor sites.

Columns correspond to different types of images: BF – blood flow, MTT – mean transit time, BV – blood volume, none – original image, HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL – 3D wavelet transforms, H – high–pass filter, L – low–pass filter.

Rows correspond to different texture parameters: HI – histogram of intensities, GLCM – Gray Level Co-occurrence Matrix, NGTDM – Neighborhood Gray Tome Difference Matrix, GLSZM – Gray Level Size Zone Matrix, 1 – mean, 2 – standard deviation, 3 – covariance, 4 – skewness, 5 – kurtosis, 6 – energy, 7 – entropy, 8 – contrast, 9 – correlation, 10 – homogeneity, 11 – variance, 12 – average, 13 – sum of entropy, 14 – sum of variance, 15 – difference entropy, 16 – difference variance, 17 – information measure of correlation 1, 18 – information measure of correlation 2, 19 – maximal correlation coefficient, 20 – coarseness, 21 – contrast, 22 – busyness, 23 – complexity, 24 – gray–level non–uniformity, 25 – size zone non–uniformity, 26 – short size emphasis, 27 – long size emphasis, 28 – low gray–level size emphasis, 29 – high gray–level size emphasis, 30 – short size low gray–level emphasis, 31 – short size high gray–level emphasis, 32 – long size low gray–level emphasis, 33 – long size high gray–level emphasis, 34 – size percentage 35 – fractal dimension.

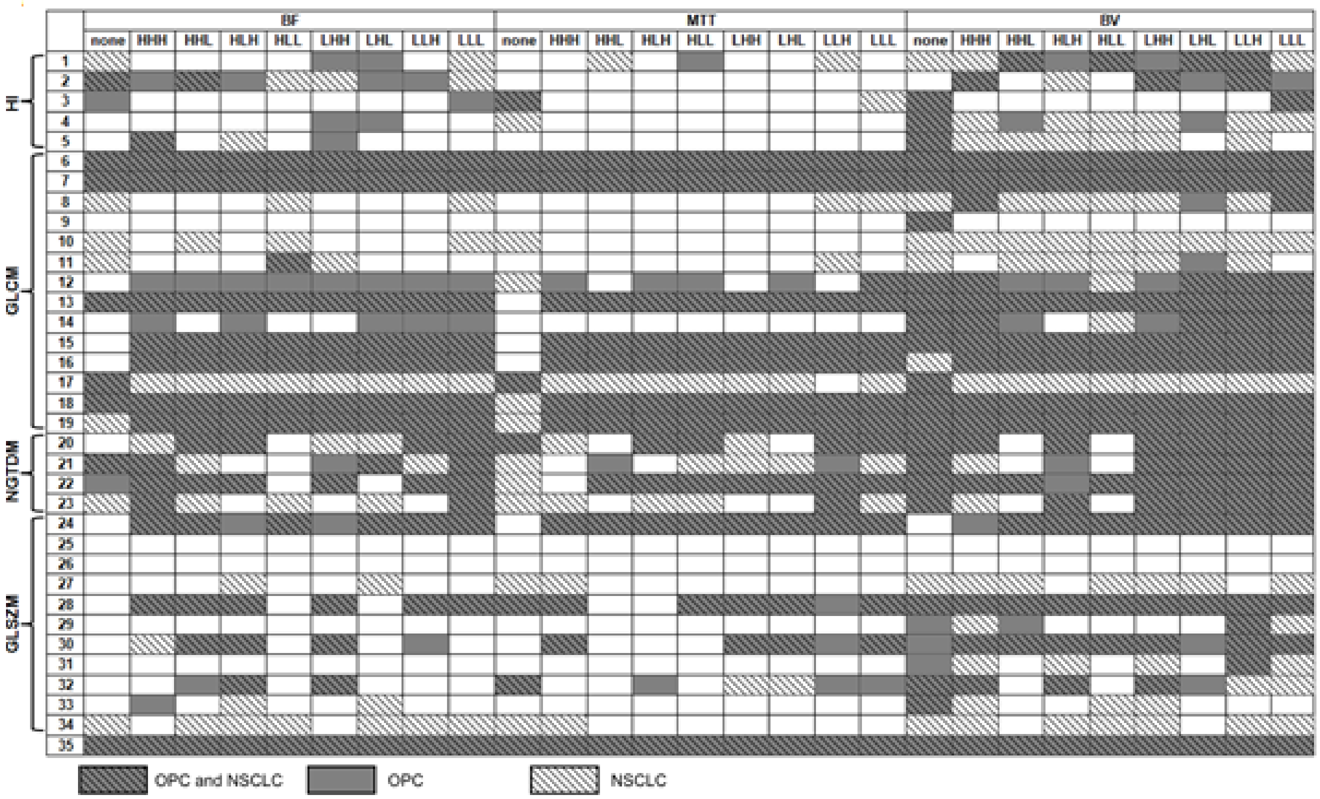


Figure S3. Robustness of texture parameters regarding noise level threshold. Changes in the noise caused instability of the mean MTT (row number 1) in both tumor sites.

Columns correspond to different types of images: BF – blood flow, MTT – mean transit time, BV – blood volume, none – original image, HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL – 3D wavelet transforms, H – high–pass filter, L – low–pass filter.

Rows correspond to different texture parameters: HI – histogram of intensities, GLCM – Gray Level Co-occurrence Matrix, NGTDM – Neighborhood Gray Tome Difference Matrix, GLSZM – Gray Level Size Zone Matrix, 1 – mean, 2 – standard deviation, 3 – covariance, 4 – skewness, 5 – kurtosis, 6 – energy, 7 – entropy, 8 – contrast, 9 – correlation, 10 – homogeneity, 11 – variance, 12 – average, 13 – sum of entropy, 14 – sum of variance, 15 – difference entropy, 16 – difference variance, 17 – information measure of correlation 1, 18 – information measure of correlation 2, 19 – maximal correlation coefficient, 20 – coarseness, 21 – contrast, 22 – busyness, 23 – complexity, 24 – gray–level non–uniformity, 25 – size zone non–uniformity, 26 – short size emphasis, 27 – long size emphasis, 28 – low gray–level size emphasis, 29 – high gray–level size emphasis, 30 – short size low gray–level emphasis, 31 – short size high gray–level emphasis, 32 – long size low gray–level emphasis, 33 – long size high gray–level emphasis, 34 – size percentage 35 – fractal dimension.

Texture analysis

Texture analysis of CT perfusion maps is more demanding in comparison to analysis of PET or CT images. The application of a HU threshold in the perfusion calculations causes discontinuities in the image, which has to be taken into account in the texture quantification. 'Not a number' value is assigned to these discontinuities in our perfusion calculation software. In this study the definitions of matrices were adjusted accordingly (see definitions below). The wavelet transform was performed in the perfusion maps using PyWavelets library and 'Coif1' wavelet. Before the transformation, discontinuities in the perfusion maps were replaced with the mean perfusion value in the tumor to avoid the loss of the outer voxels in the transformation process (Figure S4). 'Not a number' value was assigned back to the voxels in the transformed image that corresponded to the discontinuities in the original image.

The anisotropy of voxels was not take into account in the calculation of texture and wavelet transform.

The texture parameters studied in this paper are listed in Table S3. Parameters were computed on the different levels: global (histogram of intensities and fractal dimension), regional (Gray Level Size Zone Matrix) and local (Gray Level Co-occurrence Matrix and Neighborhood Gray Tome Difference Matrix).

Histogram of intensities

The parameters from the histogram of intensities were calculated in the perfusion maps before a discretization. Let X denotes the intensities of the 3D image with N voxels. $\overline{X}$ – mean of X.

$$mean = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$standard\ deviation = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\overline{X})^2}$$

$$covariance = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2}}{\overline{X}}$$

$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^2}\right)^3}$$

$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^4}{\left(\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2}\right)^4} - 3$$

Gray Level Co–occurrence Matrix

The parameters from the Gray Level Co–occurrence Matrix[1] were calculated in all 26 directions with a distance of one voxel. The final parameters were the average of all directions. If one of the voxels had a 'not a number' value the pair was not taken into account in the calculations. Let P_{ij} denotes the (i,j) entry in the Gray Level Co–occurrence Matrix, N_g – number of gray tones in a studied structure, P_{xi} – summation of the rows in the P ($P_{xi} = \sum_{j=1}^{N_g} P_{ij}$), $P_{yi} = \sum_{i=1}^{N_g} P_{ij}$, $P_{x+y}(k) = \sum_{j=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}$ where $k = i+j$, $P_{x-y}(k) = \sum_{j=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}$ where $k =\mid i-j \mid$.

$$energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}^2$$

$$contrast = \sum_{n=0}^{N_g-1} n^2 \left(\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\right) \text{, for } \mid i-j \mid = n$$

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP_{ij} - \mu_x\mu_y}{\sigma_x\sigma_y}$$

where: $\mu_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} iP_{ij}$, $\mu_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} jP_{ij}$, $\sigma_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}(i-\mu_x)^2$, $\sigma_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}(j-\mu_y)^2$.

$$\textit{sum of squares} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2 P_{ij}\text{ , where } \mu \text{ – mean of } P$$

$$\textit{inverse difference moment (homogeneity)} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{1+(i-j)^2}$$

$$\textit{sum of average} = \sum_{i=2}^{2N_g} i \cdot P_{x+y}(i)$$

$$\textit{sum of variance} = \sum_{i=2}^{2N_g}(i - \textit{sum entropy})^2 P_{x+y}(i)$$

$$\textit{sum entropy} = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log(P_{x+y}(i))$$

$$\textit{entropy} = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log(P_{ij})$$

$$\textit{difference variance} = \textit{variance of } \; p_{x-y}$$

$$\textit{difference entropy} = -\sum_{i=1}^{N_g-1} P_{x-y}(i)\log(P_{x-y}(i))$$

$$\textit{information measures of correlation 1} = \frac{\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log(P_{ij})\right)-\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log(P_{xi}P_{yj})\right)}{max\{-\sum_{i=1}^{N_g} P_{xi}\log(P_{xi}), -\sum_{j=1}^{N_g} P_{yj}\log(P_{yj})\}}$$

$$\textit{information measures of correlation 2} = \sqrt{1-\exp\left[-2\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{xi}P_{yj}\log(P_{xi}P_{yj}) + \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log(P_{ij})\right)\right]}$$

$$\textit{maximal correlation coefficient} = \sqrt{\textit{second largest eigenvalue of} \sum_{k=1}^{N_g}\frac{P_{ik}P_{jk}}{P_{xi}P_{yk}}}$$

Neighborhood Gray Tone Difference Matrix

The Neighborhood Gray Tone Difference Matrix[2] was calculated based on 26 adjacent voxels. The voxels with 'not a number' value were excluded from the average over the neighborhood region. Let s_i denotes the ith entry in the Neighborhood Gray Tone Difference Matrix, N_i – the number of voxels having gray tone i, G – number of gray tones in a studied structure, n – number of studied voxels.

$$\textit{coarseness} = \left[\varepsilon + \sum_{i=1}^{G}\frac{N_i}{n}s_i\right]^{-1}$$

where ε is a small number to prevent coarseness becoming infinite.

$$\textit{contrast} = \left[\frac{1}{N_g(N_g-1)}\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{N_i}{n}\frac{N_j}{n}(i-j)^2\right]\left[\frac{1}{n}\sum_{i=1}^{G}s_i\right]$$

where N_g is the total number of different gray levels present in the image.

$$\textit{busyness} = \frac{\sum_{i=1}^{G}\frac{N_i}{n}s_i}{\sum_{i=1}^{G}\sum_{j=1}^{G} i\frac{N_i}{n} - j\frac{N_j}{n}}$$

for $\frac{N_i}{n} \neq 0$ and $\frac{N_j}{n} \neq 0$.

$$complexity = \sum_{i=1}^{G}\sum_{j=1}^{G} \frac{(\mid i-j \mid)\left(\frac{N_i}{n}s_i + \frac{N_j}{n}s_j\right)}{N_i + N_j}$$

for $N_i \neq 0$ and $N_j \neq 0$.

Gray Level Size Zone Matrix

In the Gray Level Size Zone Matrix[3] calculations the voxels with 'not a number' value were excluded from the analysis. Let P_{ij} denotes the (i,j) entry in the Gray Level Size Zone Matrix, i – gray value, j – size, n_r – number of homogeneous areas inside a studied structure.

$$gray\text{–}level\ non\text{–}uniformity = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} P_{ij}\right)^2$$

$$size\ zone\ non\text{–}uniformity = \frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M} P_{ij}\right)^2$$

$$short\ size\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{j^2}$$

$$long\ size\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot j^2$$

$$low\ gray\text{–}level\ size\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2}$$

$$high\ gray\text{–}level\ size\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2$$

$$short\ size\ low\ gray\text{–}level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2\cdot j^2}$$

$$short\ size\ high\ gray\text{–}level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot i^2}{j^2}$$

$$long\ size\ low\ gray\text{–}level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot j^2}{i^2}$$

$$long\ size\ high\ gray\text{–}level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2\cdot j^2$$

$$size\ percentage = \frac{n_r}{number\ of\ voxels\ in\ the\ studied\ structure}$$

Fractal dimension

The fractal dimension was calculated using box contouring technique and fixed grid scans excluding the voxels with 'not a number' value.[4]

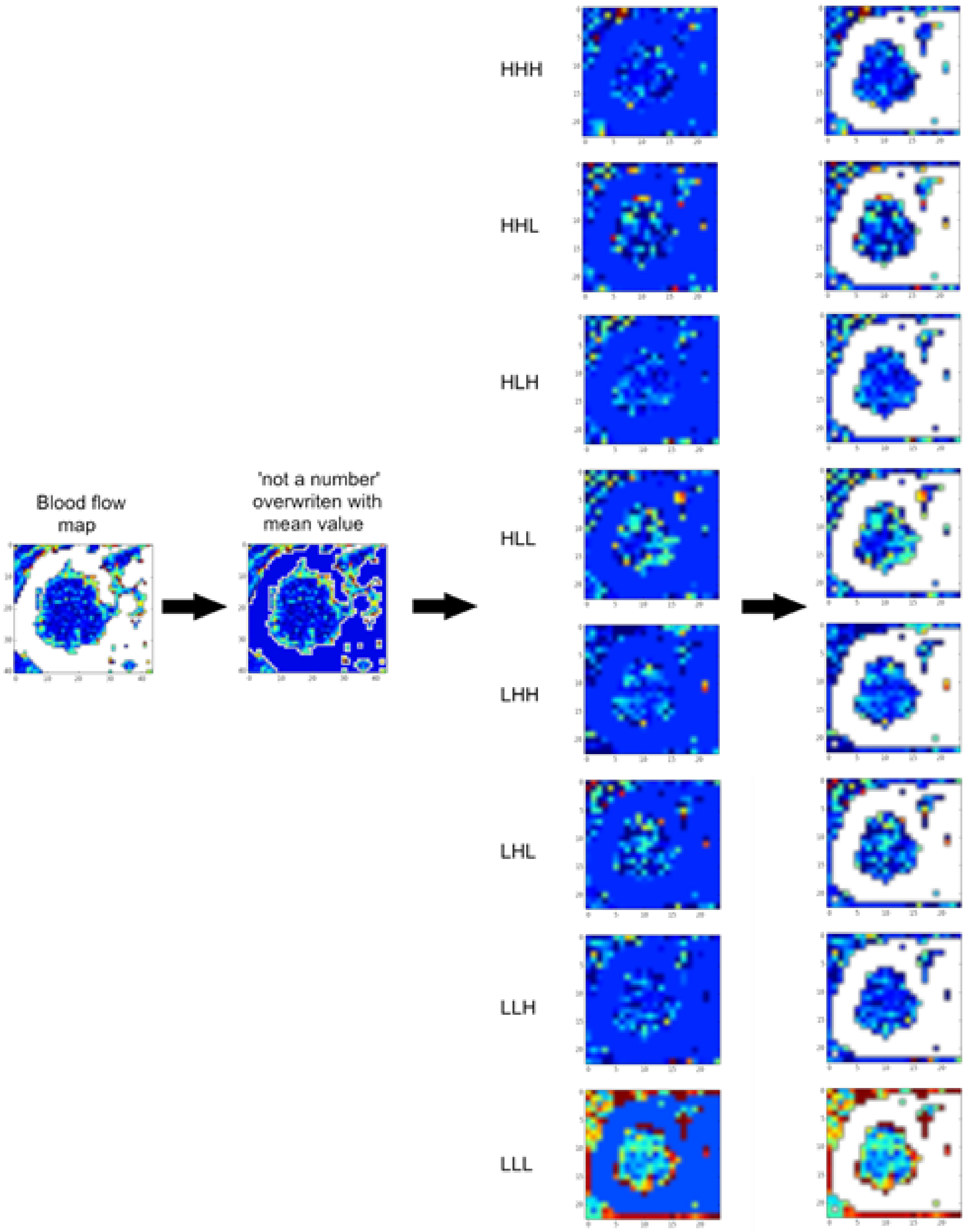


Figure S4. The scheme of wavelet transform in CT perfusion maps, H – high–pass filter, L – low–pass filter.

Model	**Level**	**Parameter**
Statistical model **Histogram of intensities**	Global	Mean
		Standard deviation
		Covariance
		Skewness
		Kurtosis
Statistical model **Gray Level Co–occurrence Matrix**	Local	Energy
		Contrast
		Correlation
		Sum of squares
		Inverse difference moment (homogeneity)
		Sum of average
		Sum of variance
		Sum entropy
		Entropy
		Difference variance
		Difference entropy
		Information measures of correlation 1
		Information measures of correlation 2
		Maximal correlation coefficient
Statistical model **Neighborhood Gray Tone Difference Matrix**	Local	Coarseness
		Contrast
		Busyness
		Complexity
Statistical model **Gray Level Size Zone Matrix**	Regional	Gray–level non-uniformity
		Size zone non–uniformity
		Short size emphasis
		Long size emphasis
		Low gray–level size emphasis
		High gray–level size emphasis
		Short size low gray–level emphasis
		Short size high gray–level emphasis
		Long size low gray–level emphasis
		Long size high gray–level emphasis
		Size percentage
Model–based	Global	Fractal dimension

Table S3. List of the computed texture parameters.

References

1. R.M. Haralick, K. Shanmugam, and I. Dinstein, ”Textural features for image classification,” IEEE Trans Syst Man Cybern. **3**, 610-621 (1973).
2. M. Amadasun, and R. King, ”Texural features corresponding to textural properties,” IEEE Trans Syst Man Cybern. **19**, 1264-1274 (1989).
3. G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N, Levy, J. Sequeira, and J.L. Mari, ”Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognition Inf Process,” **2009**, 140-145 (2009).
4. A. Napolitano, S. Ungania, and V. Cannata. ”Fractal Dimension Estimation Methods for Biomedical Images” in *MATLAB - A Fundamental Tool for Scientific Computing and Engineering Applications – Vol. 3*, edited by V.N. Katsikis (InTech, 2012).

3.5 CT radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma.

Bogowicz Marta[1], Riesterer Oliver[1], Ikenberg Kristian[2], Stieb Sonja[1], Moch Hogler[2], Studer Gabriela[1], Guckenberger Matthias[1], Tanadini-Lang Stephanie[1]

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.
[2] Institute of Clinical Pathology, University Hospital Zurich and University of Zurich, Zurich Switzerland.

Status of the manuscript:

published Int J Rad Onc Biol Phys. 2017, in press

Author contribution Marta Bogowicz:

- Modification of image segmentation
- Full data analysis: radiomics computation and training of a local tumor control prognostic model and the HPV status prediction model (results for Figures 1-2 and Tables 2-4)
- Preparation of all supplementary files
- Manuscript writing, revision, editing and proofreading.

International Journal of
Radiation Oncology
biology ● physics

www.redjournal.org

Physics Contribution

Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma

Marta Bogowicz, MSc,[*] Oliver Riesterer, MD,[*] Kristian Ikenberg, MD,[†] Sonja Stieb, MD,[*] Holger Moch, MD,[†] Gabriela Studer, MD,[*] Matthias Guckenberger, MD,[*] and Stephanie Tanadini-Lang, PhD[*]

Departments of [*]Radiation Oncology, and [†]Pathology and Molecular Pathology, University Hospital Zurich and University of Zurich, Zurich, Switzerland

Received Dec 22, 2016, and in revised form May 15, 2017. Accepted for publication Jun 5, 2017.

Summary

This study shows that higher pretreatment tumor heterogeneity, quantified using radiomics of contrast-enhanced computed tomography, is associated with worse prognosis in head and neck squamous cell carcinoma and the absence of human papillomavirus infection. Overall, our analysis provides evidence that a radiomic signature might be a useful input in risk assessment in addition to clinical parameters (tumor stage, tumor volume, and HPV status).

Purpose: This study aimed to predict local tumor control (LC) after radiochemotherapy of head and neck squamous cell carcinoma (HNSCC) and human papillomavirus (HPV) status using computed tomography (CT) radiomics.

Methods and Materials: HNSCC patients treated with definitive radiochemotherapy were included in the retrospective study approved by the local ethical commission (93 and 56 patients in the training and validation cohorts, respectively). Three hundred seventeen CT radiomic features, including those based on shape, intensity, texture, and wavelet transform, were calculated in the primary tumor region. Cox and logistic regression models were built to predict LC and HPV status, respectively. The best-performing features in the univariable analysis were included in the multivariable analysis after the exclusion of redundant features. The quality of the models was assessed using the concordance index (CI) for modeling of LC and receiver operating characteristics area under the curve (AUC) for HPV status prediction. The radiomics LC model was compared to a model incorporating clinical parameters (tumor stage, volume, and HPV status) and a mixed model.

Results: A radiomic signature comprising 3 features was significantly associated with LC ($CI_{training} = 0.75$ and $CI_{validation} = 0.78$), showing that tumors with a more heterogeneous CT density distribution are at risk for decreased LC. The addition of clinical parameters to the radiomics model slightly improved the model in the training cohort but not in the validation cohort. Another radiomic signature showed good

---

Reprint requests to: Marta Bogowicz, MSc, Department of Radiation Oncology, University Hospital Zurich, Rämistrasse 100, Zurich 8091, Switzerland. Tel: (+41) 44-255-4116; E-mail: marta.nesteruk@usz.ch

Supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant from the Matching Fund of the University of Zurich, and by a research grant from Merck (Schweiz) AG.

Conflict of interest: none.

Supplementary material for this article can be found online at www.redjournal.org.

Int J Radiation Oncol Biol Phys, Vol. ■, No. ■, pp. 1−8, 2017

http://dx.doi.org/10.1016/j.ijrobp.2017.06.002

performance in HPV status prediction ($\text{AUC}_{\text{training}} = 0.85$ and $\text{AUC}_{\text{validation}} = 0.78$) and indicated that HPV-positive tumors have a more homogenous CT density distribution.

Conclusions: Heterogeneity of HNSCC tumor density, quantified by CT radiomics, is associated with LC after radiochemotherapy and HPV status.

Introduction

Head and neck squamous cell carcinoma (HNSCC) is the fifth most common cancer worldwide. The standard of care for advanced unresectable cases is definitive radiochemotherapy. The clinical outcome after radiochemotherapy is very heterogeneous, ranging from about 80% locoregional control and 5 years overall survival in human papillomavirus (HPV)-positive oropharyngeal carcinoma (OPC) to below 50% in HPV-negative OPC and non-OPC (1, 2). In all cases, distant metastasis is infrequent and is mainly a consequence of locoregional treatment failure. Therefore, improvement of radiochemotherapy is essential. An important approach is the development of tumor biomarkers that can predict locoregional tumor control to identify patients for treatment intensification.

Radiomics is a promising tool for noninvasive characterization of tumor phenotype. It extracts a large number of quantitative features from radiographic images: regions of interest are characterized regarding their shape, intensities, and texture of both spatial and frequency domains (for example, using a wavelet transform) (3). A correlation between radiomic features and treatment response has been shown using different image modalities and for different tumor types (4, 5). To provide a biological interpretation of radiomics, several groups have tried to relate a radiomic signature (set of radiomic parameters) to differential gene expressions, including tumor proliferation and molecular subtypes (6, 7).

Although several computed tomography (CT) radiomics-based prognostic models have been proposed for HNSCC (8), it is not fully understood why a particular radiomic phenotype is associated with treatment response. CT radiomics of the primary tumor was successfully used to stratify patients into low-risk and high-risk groups in terms of patient overall survival (6, 8-10). It was shown to further improve survival models based on clinical parameters, such as TNM stage and tumor volume (6). Whereas overall survival is an important clinical endpoint, it is also influenced by non−tumor-related factors (11), which cannot be captured by pretreatment radiomics of the primary tumor. However, especially in the case of HNSCC, it is important to understand the influence of intratumor heterogeneity on the local response to treatment. Therefore, a study using local tumor control as endpoint is of high clinical interest. Furthermore, the correlation of radiomic phenotype to the important outcome prognostic factor HPV is of interest and links image analysis with tumor biology (12, 13). Exploratory studies have already indicated a relationship between HPV infection and the heterogeneity of imaging-based tumor density (12, 13). However, these studies were based on small patient numbers and lacked validation cohorts.

In this study we investigated whether a CT-based radiomic phenotype of HNSCC is correlated with local tumor control after radiochemotherapy. Furthermore, using 2 large patient cohorts, we studied the influence of HPV infection on tumor radiomics.

Methods and Materials

Studied population

This study was a retrospective analysis approved by the local ethical commission. The training cohort consisted of 93 patients with stage III and IV HNSCC who were treated between 2003 and 2013 with definitive radiation therapy (on average 70 Gy) either in combination with cisplatin (40 mg/m^2, up to 7 cycles) or cetuximab (loading dose 400 mg/m^2 followed by 250 mg/m^2 weekly). Induction chemotherapy or prior surgery (biopsy allowed), were exclusion criteria. Additionally, data from 56 patients with the same tumor stages from an institutional phase 3 prospective study (NCT01435252) with a standardized planning CT imaging protocol were selected as a validation cohort. In the latter study, patients were treated with triple therapy consisting of radiation therapy and weekly concurrent cisplatin/cetuximab (same doses as in training cohort) with or without consolidation cetuximab (500 mg/m^2 biweekly × 6). Radiation therapy was delivered with an intensity modulated technique in all patients.

All patients underwent pretreatment contrast-enhanced planning CT imaging. Patient characteristics and imaging protocols are presented in Table 1. In all patients, the HPV status was determined by immunohistochemical p16 staining. Local recurrence was defined as biopsy-proven recurrence within the high-dose planning target volume of the primary tumor. Follow-up consisted of examinations every 2 to 3 months by ear, nose, and throat specialists and repeated 18F-fluorodeoxyglucose positron emission tomography 18F-FDG PET/CT imaging.

Table 1 Studied patient cohorts and imaging protocol details

Characteristic	Training cohort	Validation cohort
Number of patients	93	56
Number of recurrences	24 (25%)	14 (25%)
Median follow-up [months]	47 (3-156)*	16 (3-28)*
Age [years]*	61.9 (41.9-74.5)	59.8 (47.2-75.7)
HPV status		
p16 positive	36 (39%)	26 (46%)
p16 negative	57 (61%)	30 (54%)
Tumor stage		
T1/T2	33 (36%)	9 (16%)
T3/T4	60 (64%)	47 (84%)
Nodal stage		
N0	15 (16%)	10 (18%)
N1	5 (5%)	4 (7%)
N2	69 (74%)	40 (71%)
N3	4 (5%)	2 (4%)
Tumor site		
Oropharynx	70 (75%)	30 (54%)
Hypopharynx	18 (19%)	8 (14%)
Larynx	4 (5%)	9 (16%)
Oral cavity	1 (1%)	9 (16%)
Sex		
Female	21 (23%)	13 (23%)
Male	72 (77%)	43 (77%)
Scanning parameters		
CT scanners	Siemens SOMATOM Volume Zoom (n=64) Siemens SOMATOM Definition AS (n=22) GE Discovery STE (n=5)	Siemens SOMATOM Definition AS (n=56)
Slice thickness [mm]	1.25-3.30	2
In-plane resolution [mm]	0.85-1.95	0.98-1.56
kV	120-140	120
mAs	60-450	183-450

* Median; range in parentheses.

Radiomics analysis and image postprocessing

An in-house-developed radiomics software implementation written in Python programming language was used. It provides 3-dimensional (3D) analysis of medical images using all 4 feature extraction methods:

1. Shape (n=11)
2. Intensity (n=5)
3. Texture: the Gray Level Co-occurrence matrix (n=14), the Neighborhood Gray Tone Difference matrix (n=4) and the Gray Level Size Zone matrix (n=11)
4. Wavelet transform using Coiflet function (8 sub-bands, n=272)

In total, 317 features were extracted per patient. For intensity, texture, and wavelet analysis, images were resized to the symmetrical voxels of size 3.3 mm, corresponding to the lowest image resolution (sagittal) in the training cohort. Additionally, a Hounsfield Units range of −20 to 180 HU was applied to limit the analysis to soft tissue only. To quantify the texture, images were resampled to equally spaced bins of 5 HU. A detailed description of radiomics implementation is presented in the Appendix (available online at www.redjournal.org).

The tumor heterogeneity was studied inside the primary gross tumor volume (GTV) defined on the contrast-enhanced planning CT by a radiation oncologist with more than 10 years of experience. Contours were postprocessed to account for metal artifacts occurring in the CT images. Contours were removed from artifact-affected slices. Scans with more than a half of the contoured slices affected by metal artifacts were not included in the analysis.

Radiomic interfeatures correlations

Radiomic features were grouped according to principal component analysis using data from the training cohort to account for the interfeatures correlations. The Horn method was used to choose the optimal number of retained components (14). The contribution of radiomic features to principal components was determined. Features were assigned to a principal component to which they contributed the most. Each principal component corresponded to one group of correlated features. The statistical analysis was performed in R (version 3.2.4) (15).

Prognostic value of radiomics for local tumor control

Radiomic features prognostic for local tumor control were preselected in the univariable Cox regression analysis. The results were corrected for multiple testing with a false discovery rate (FDR) <0.20. Additionally, the correlated features were excluded from further analysis by selecting only the most prognostic parameter per principal component feature group. Feature prognostic value was quantified using a concordance index (CI) (16). The preselected features were entered in the multivariable Cox regression analysis. The radiomics-based model was compared with a multivariable model created with clinical parameters: tumor volume, T stage (T1/T2 vs T3/T4), and HPV status. Finally, a mixed model was created, wherein both radiomic features and clinical parameters were incorporated in the multivariable Cox model. All multivariable models were built using an Akaike information criterion in the backward selection of variables. The prognostic value of obtained models was compared based on the CI, using the Wilcoxon test ($P<.05$) and the bootstrap method with 100 randomly selected samples to calculate the CI distribution. The results were verified in the validation cohort. Curves were split based on the optimal threshold from the receiver operating characteristic curve at 18 months maximizing both sensitivity and specificity and compared using G-rho test ($P<.05$).

Prediction of HPV status

The radiomic features for HPV status prediction were also preselected using principal component analysis in combination with univariable logistic regression. The univariable model accuracy was quantified by the area under the receiver operating characteristic curve (AUC). The results were corrected for multiple testing with FDR <0.20. A radiomic feature characterized by the highest AUC was selected in each principal component group. This set was later used in multivariable logistic regression with backward selection of the variables. The cutoff threshold of the final model was set to maximize both sensitivity and specificity in the training cohort. The results were again verified in the validation cohort.

Results

Radiomics-based prognostic model for local tumor control

Nine components were selected in the principal component analysis, which led to 9 groups of correlated radiomic features. The group sizes varied from 21 to 66 features. In the univariable Cox regression analysis, 83 of 317 radiomic features were found to be prognostic for local tumor control (FDR <0.20). The list of the most prognostic features, accounting for interfeature correlations, is shown in Table 2. In the multivariable analysis with the backward selection of variables, 3 out of 9 previously mentioned features were selected (HHH large zone high gray-level emphasis, LLL sum entropy, and LLH difference variance), resulting in a model with CI of 0.75. This model was found to be prognostic also in the validation cohort with CI = 0.78 and significant for stratification into groups with low risk and high risk of recurrence (Fig. 1, Table 3). Additionally, the model was prognostic in the subgroup analysis of OPC and of HPV-negative tumors (Table E1 and Fig. E2 of the Appendix; available online at www.redjournal.org). Different CT scanners and kV settings did not show a major influence on the radiomic features included in the model (Fig. E6; available online at www.redjournal.org).

Comparison of the radiomics-based and clinical-based prognostic models for local tumor control

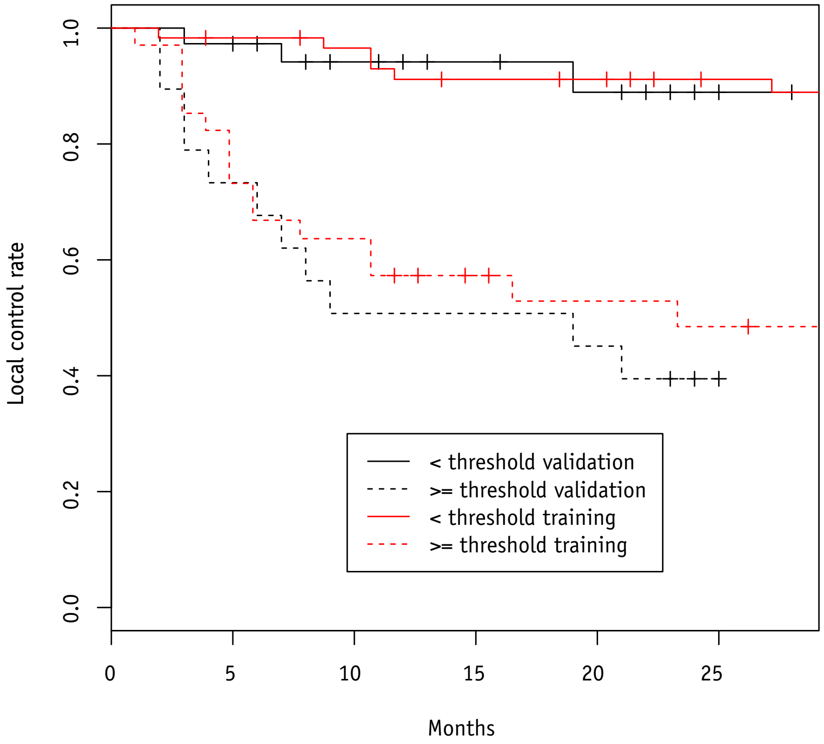
The radiomics-based model was compared with a model incorporating known prognostic clinical parameters (tumor volume, T stage, and HPV status) and also with a mixed model (radiomics and clinical parameters). The parameters chosen in the multivariable analysis and the results of Wilcoxon test, studying differences between the models, are presented in Table 3. The combination of radiomics (LLH difference variance) and clinical parameters (HPV and tumor volume) was found to have the highest prognostic power (CI = 0.80) in the training cohort but performed worse than the radiomics-based model in the validation cohort. The same trend was observed for the subgroup of OPC (Table E1 of the Appendix; available online at www.redjournal.org). In the mixed model, the

Table 2 Uncorrelated (independent) radiomic features prognostic for local tumor control in the univariable Cox regression analysis

Feature group	Feature name	CI	FDR
1	LLL sum entropy*	0.75	0.11
2	LHL IMC1	0.67	0.12
3	HHH large zone high gray-level emphasis*	0.68	0.13
4	HHL contrast$_{NGTDM}$	0.69	0.19
5	LLH difference variance*	0.65	0.16
6	HLH complexity	0.64	0.16
7	Sum entropy	0.64	0.16
8	LHH MCC	0.68	0.12
9	LHL low gray-level emphasis	0.70	0.12

Abbreviations: CI = concordance index; FDR = false discovery rate.

* The features selected in the multivariable Cox regression analysis.



59	57	54	47	46	42	< threshold training
34	27	20	15	12	11	>= threshold training
37	34	27	22	17	3	< threshold validation
19	13	9	9	8	1	>= threshold validation

Fig. 1. Radiomics-based local tumor control model. The model consists of 3 radiomic features: HHH large size high gray-level emphasis, LLL sum entropy, and LLH difference variance. Survival curves split significantly (G-rho test $P < .001$) in both training and validation cohorts based on the optimal sensitivity-specificity threshold.

single radiomic feature was not significant; however, it had a significant impact on the improvement of the model discriminative power (Table E2 of the Appendix; available online at www.redjournal.org).

Prediction of HPV status

In total, 86 radiomic features were significant for HPV status prediction (FDR <0.20). Per group of correlated

Table 3 Comparison of the prognostic models for local tumor control

Model	Parameters of the model	Training cohort CI	Training cohort *P*	Validation cohort CI	Validation cohort *P*
Radiomics	HHH large zone high gray-level emphasis LLL sum entropy LLH difference variance	0.75 (0.74-0.76)	<.001	0.78 (0.77-0.79)	<.001
Clinical	Stage Tumor volume HPV	0.79 (0.78-0.80)		0.73 (0.71-0.75)	
Radiomics	HHH large zone high gray-level emphasis LLL sum entropy LLH difference variance	0.75 (0.74-0.76)	<.001	0.78 (0.77-0.79)	<.046
Radiomics + clinical	LLH difference variance Tumor volume HPV	0.80 (0.79-0.81)		0.76 (0.75-0.77)	

Abbreviation: CI = concordance index (its 95% confidence interval in parentheses).
Wilcoxon test *P* are given for the comparison of the models performance (radiomics vs clinical and radiomics vs radiomics + clinical).

Table 4 Uncorrelated (independent) radiomic features for HPV status prediction

Feature group	Feature name	AUC	FDR
1	LLL gray-level non-uniformity	0.77	0.053
2	LLL standard deviation*	0.83	0.002
3	Difference entropy	0.76	0.004
4	LLL small zone high gray-level emphasis*	0.78	0.004
5	HHL difference entropy*	0.67	0.143
6	Compactness 2	0.66	0.077
7	Standard deviation	0.79	0.004
8	Coefficient of variation*	0.69	0.071
9	Spherical disproportion	0.66	0.069

Abbreviations: AUC = area under the curve; FDR = false discovery rate.

* The features selected in the multivariable logistic regression.

features, the feature characterized with the highest AUC was chosen (Table 4) and entered in the multivariable logistic regression. The final model comprised 4 radiomic features:

1. LLL standard deviation
2. LLL small zone high gray-level emphasis
3. HHL difference entropy
4. Coefficient of variation

The final radiomic model for HPV status prediction showed a good performance in both the training cohort (AUC = 0.85) and the validation cohort (AUC = 0.78). The specificity of the model was higher in the validation cohort than in the training cohort ($\text{specificity}_{\text{training}} = 0.72$ and $\text{specificity}_{\text{validation}} = 0.82$), whereas the sensitivity was lower ($\text{sensitivity}_{\text{training}} = 0.72$ and $\text{sensitivity}_{\text{training}} = 0.62$). Figure 2 presents the receiver operating characteristics for the model. The HPV radiomic signature was independent from the local control signature because the radiomic features in these 2 models were not correlated and the HPV prediction model was valid in patient subgroups regardless of tumor control (Appendix Fig. E3). Similar to the local tumor control model, we did not observe major variations in our HPV radiomic signature depending on the CT scanner and kV settings (Fig. E7 of the Appendix; available online at www.redjournal.org).

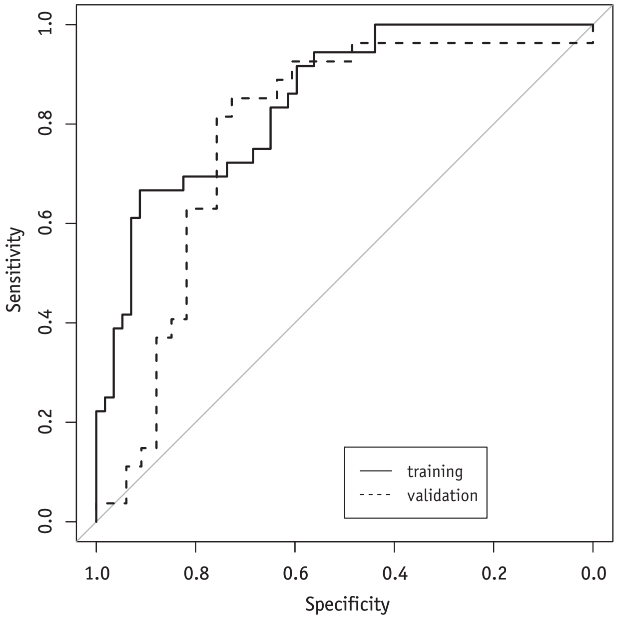


Fig. 2. Receiver operating characteristics for HPV status prediction using radiomics with area under the curve (AUC): AUC training = 0.85, AUC validation = 0.78.

Discussion

This study evaluated the prognostic value of contrast-enhanced CT-based radiomics in HNSCC regarding local tumor control and HPV status. The radiomics-based local control model was compared with a model based on prognostic clinical parameters (tumor stage, tumor volume, and HPV status) and with the mixed model (clinical plus radiomics). The performance of the models was verified in an independent cohort of patients. Although the radiomics-based model achieved a lower concordance index in comparison with the mixed model in the training cohort, it showed a significantly higher performance in the validation cohort. The stable performance of the radiomics-based model in both the training and the validation patient cohorts emphasizes its high credibility. Radiomics was also shown to be closely associated with the HPV status with AUC values of 0.85 and 0.78 in the training cohort and the validation cohort, respectively.

In the case of HNSCC, locoregional control closely correlates with overall survival (17). Prediction of local treatment efficacy is of high interest in modifying treatment according to the individual tumor's radiosensitivity. The radiomics model created by us was highly prognostic for local tumor control; the addition of the clinically relevant biomarker, HPV status, could only slightly improve the model in the training cohort but not in the validation cohort. Similar results were observed for the analysis of OPC patients only (Table E1 of the Appendix; available online at www.redjournal.org), which indicates slight overfitting of the mixed model (radiomics and clinical predictors) in the training cohort. The value of the radiomic feature (LLH difference variance) in the mixed model should be studied in more detail. This covariate was not significant in the multivariable modeling but had a positive impact on the discriminative power of the model, especially in the validation cohort. The training and validation cohorts were not perfectly matched. The validation cohort was characterized by a higher percentage of more advanced tumors, lower share of OPC, and higher percentage of HPV-positive tumors. Additionally, the follow-up times were longer in the training cohort; however, a similar percentage of tumor recurrences was observed in both cohorts. Overall, we showed that a radiomic signature might be a useful input in risk assessment in addition to clinical parameters. An additional validation could confirm or contradict the benefit of including radiomics in the modeling.

In our study, we observed that local tumor control is linked more closely to heterogeneity of the CT density than to tumor shape, including tumor volume. Tumors that are characterized by more homogenous CT density have a higher probability of local control (Fig. E4; available online at www.redjournal.org). The HHH large zone high gray-level emphasis was lower in the group of controlled tumors, which corresponds to fewer patches of large coefficients in the image transformed with 3D high-pass filter. Additionally, the LLL sum entropy was also lower, which relates to a more repeatable pattern of intensity changes in the neighboring voxels. The biology behind this more homogenous texture is unknown but warrants further investigation. It can only be speculated that texture heterogeneity might be related to differential biology within the tumor. On the biological level, a more heterogeneous microenvironment (eg, related to tumor perfusion, necrosis, or tumor hypoxia) is also known to be associated with a more resistant phenotype, especially in HNSCC. Our findings agree with the findings of a previously published study showing that tumor heterogeneity in CT radiomics was related to worse overall survival (6). Our model shows a higher performance than the overall survival model, which suggest that an additional input, for example the radiomics of lymph node metastasis, might improve the latter.

Another set of CT radiomic features was found to be associated with p16 staining. The integration in the host genome and the expression of the viral protein E7, through inactivation of the retinoblastoma protein, leads to an upregulation of p16^{INK4a} and correlates strongly with active HPV infection. Therefore, the immunohistochemical detection of the expression level of p16^{INK4a} is now a widely accepted surrogate marker for transcriptionally active HPV infections (18). Our results show that HPV-positive tumors seem to be more homogenous in CT density (Fig. E5 of the Appendix; available online at www.redjournal.org). The HPV-positive tumors were characterized by a significantly lower coefficient of variation and LLL standard deviation, which refers to a smaller width of density distribution in the tumor. They also presented lower LLL small zone high gray-level emphasis, corresponding to a lower number of small regions with high intensity. The more homogenous density of HPV-positive tumors also seems to fit the hypothesis that tumor homogeneity is correlated to a more favorable phenotype of HNSCC. Our results, observed in 2 independent patient cohorts, support the hypothesis presented in the explanatory studies (12, 13) that HPV infection is related to tumor density homogeneity.

Although both local tumor recurrence and HPV radiomic signatures refer to heterogeneity in tumor CT density, they consisted of different radiomic features. No strong correlation between features in different models was observed. The local recurrence signature was also prognostic in the subgroup of HPV-negative patients (Fig. E2; available online at www.redjournal.org). It suggests that the local recurrence radiomic signature is not driven by HPV status.

Certain limitations were present in this study. The analysis was based on single-center data, although patients enrolled in this study were divided into training and validation cohorts; also, the scanning protocol was not standardized for the training cohort, and 3 different types of CT scanners were used. The proposed models should be further validated in an external cohort to test them against different contouring protocols and different scanners. The interobserver variability in GTV definition in head and neck cancer is a known issue (19); however, its influence on CT radiomics remains unclear. Additionally, the stability of the features should be evaluated in a test-retest study. Furthermore, patients in the training and validation cohorts were treated with different forms of concurrent systemic therapy: with either cisplatin or cetuximab weekly in the training cohort and with both agents, cisplatin and cetuximab, in the consolidation cohort. Of note, the radiomics model seemed to be robust against the choice of concurrent systemic therapy. Moreover, the scanning grid size varied among the patients, which is known to influence stability of texture (20). However, there is no consensus in the literature about the best correction method; in our study the resampling approach to isotropic voxels using linear interpolation was used. These results cannot be directly translated to any other resolution. The investigation of HPV influence on the tumor structure in CT images is just a first step in understanding tumor heterogeneity in CT. Further studies correlating CT radiomics to histology markers and gene expression are needed.

In conclusion, this study shows that tumor heterogeneity, quantified using contrast-enhanced CT radiomics, is associated with local tumor control and HPV status in HNSCC. The radiomics-based local tumor control model showed a better performance than did models incorporating clinical parameters in our validation cohort. However, this result should be further investigated on a larger, external validation set to provide more evidence on the prognostic power of CT radiomics.

References

1. Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010;363:24-35.
2. Lassen P, Primdahl H, Johansen J, et al. Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. *Radiother Oncol* 2014;113:310-316.
3. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441-446.
4. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol* 2016;61:150-166.
5. Hatt M, Tixier F, Pierce L, et al. Characterization of PET/CT images using texture analysis: The past, the present… any future? *Eur J Nucl Med Mol Imaging* 2017;44:151-165.
6. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.

7. Yang D, Rao G, Martinez J, et al. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. *Med Phys* 2015;42: 6725-6735.
8. Wong AJ, Kanwar A, Mohamed AS, et al. Radiomics in head and neck cancer: From exploration to application. *Transl Cancer Res* 2016;5: 371-382.
9. Leijenaar R, Carvalho S, Hoebers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol* 2015;54:1423-1429.
10. Parmar C, Leijenaar R, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Sci Rep* 2015;5:11004.
11. Harrison LB, Sessions RB, Hong WK. Head and Neck Cancer: A Multidisciplinary Approach. Philadelphia: Lippincott Williams & Wilkins, Wolters Kluwer; 2009.
12. Buch K, Fujita A, Li B. Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on CT. *AJNR Am J Neuroradiol* 2015;36:1343-1348.
13. Fujita A, Buch K, Li B, et al. Difference between HPV-positive and HPV-negative non-oropharyngeal head and neck cancer: Texture analysis features on CT. *J Comput Assist Tomogr* 2016;40:43-47.
14. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179-185.
15. R Development Core Team. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.
16. Harrell FE Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2011.
17. Michiels S, Le Maître A, Buyse M, et al. Surrogate endpoints for overall survival in locally advanced head and neck cancer: Meta-analyses of individual patient data. *Lancet Oncol* 2009;10: 341-350.
18. Krupar R, Hartl M, Wirsching K, et al. Comparison of HPV prevalence in HNSCC patients with regard to regional and socioeconomic factors. *Eur Arch Otorhinolaryngol* 2014;271:1737-1745.
19. Riegel AC, Berson AM, Destian S, et al. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int J Radiat Oncol Biol Phys* 2006;65:726-732.
20. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 2016;6: 234283.

Appendix

CT radiomics predicts HPV status and local tumor control after radiochemotherapy in head and neck squamous cell carcinoma

Marta Bogowicz, Oliver Riesterer, Kristian Ikenberg, Sonja Stieb, Holger Moch, Gabriela Studer, Matthias Guckenberger, Stephanie Tanadini-Lang

Radiomics

The following approaches were used to determine radiomics: histogram of intensities analysis, texture analysis (the Gray Level Co-occurrence Matrix, the Neighborhood Gray Tome Difference Matrix and the Gray Level Size Zone Matrix), shape analysis and wavelet transform analysis.

Histogram of intensities

The parameters from the histogram of intensities were calculated in the perfusion maps before a discretization. Let X denotes the intensities of the 3D image with N voxels. $\bar{X}$ - mean of X.

$$mean = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$standard\ devation = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

$$coefficient\ of\ variation = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}}{\bar{X}}$$

$$skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}\right)^3}$$

$$kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}\right)^4} - 3$$

The Gray Level Co-occurrence Matrix

The parameters from the Gray Level Co-occurrence Matrix [1] were calculated in all 26 directions with a distance of one voxel. The final parameters were the average of all directions. If one of the voxels had a 'not a number' value the pair was not taken into account in the calculations. Let P_{ij} denotes the (i,j) entry in the Gray Level Co-occurrence Matrix, N_g - number of gray tones in a studied structure, $P_{xi} = \sum_{j=1}^{N_g} P_{ij}$, $P_{yj} = \sum_{i=1}^{N_g} P_{ij}$, $P_{x+y}(k) = \sum_{j=1}^{N_g}\sum_{i=1}^{N_g} P_{ij}$, where $k = i + j$, $P_{x-y}(k) = \sum_{j=1}^{N_g}\sum_{i=1}^{N_g} P_{ij}$, where $k = |i - j|$.

$$energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}^2$$

$$contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-j)^2 P_{ij}$$

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP_{ij} - \mu_x\mu_y}{\sigma_x\sigma_y}$$

where: $\mu_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} iP_{ij}$, $\mu_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} jP_{ij}$, $\sigma_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}(i-\mu_x)^2$, $\sigma_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\left(j-\mu_y\right)^2$

$sum\ of\ squares = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2 P_{ij}$, where μ – mean of P

$$inverese\ difference\ moment\ (homogeneity) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P_{ij}}{1+(i-j)^2}$$

$$sum\ of\ average = \sum_{i=2}^{2N_g} i \cdot P_{x+y}(i)$$

$$sum\ of\ varaince = \sum_{i=2}^{2N_g}(i - sum\ of\ average)^2 P_{x+y}(i)$$

$$sum\ entropy = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log_2 P_{x+y}(i)$$

$$entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij}$$

$$difference\ variance = variance\ of\ P_{x-y}$$

$$difference\ entropy = -\sum_{i=0}^{N_g-1} P_{x-y}(i)\log_2 P_{x-y}(i)$$

$$information\ measure\ of\ correlation\ 1 = \frac{-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij} - \left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right)}{max\left\{\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{xi}\log_2 P_{xi}\right),\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{yj}\log_2 P_{yj}\right)\right\}}$$

$$information\ measure\ of\ correlation\ 2 = \sqrt{1-\exp\left[-2-\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right)+\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij}\right]}$$

$$maximal\ correlation\ coefficient = \sqrt{\text{second largest eigenvalue of } \sum_{k=1}^{N_g}\frac{P_{ik}P_{jk}}{P_{xi}P_{ji}}}$$

The Neighborhood Gray Tone Difference Matrix

The Neighborhood Gray Tone Difference Matrix [2] was calculated based on 26 adjacent voxels. The voxels with ‘not a number’ value were excluded from the average over the neighborhood region. Let s_i

denotes the i^{th} entry in the Neighborhood Gray Tone Difference Matrix, N_i - the number of voxels having gray tone i, G - number of gray tones in a studied structure, n - number of studied voxels.

$$coarseness = \left[\epsilon + \sum_{i=1}^{G} \frac{N_i}{n} s_i\right]^{-1}$$

where ϵ is a small number to prevent coarseness becoming infinite.

$$contrast = \left[\frac{1}{N_g(N_g - 1)} \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{N_i}{n} \frac{N_j}{n} (i - j)^2\right] \left[\frac{1}{n} \sum_{i=1}^{G} s_i\right]$$

where N_g is the total number of different gray levels present in the image.

$$busyness = \frac{\sum_{i=1}^{G} \frac{N_i}{n} s_i}{\sum_{i=1}^{G} \sum_{j=1}^{G} i \frac{N_i}{n} - j \frac{N_j}{n}}$$

for $\frac{N_i}{n} \neq 0$ and $\frac{N_j}{n} \neq 0$

$$complexity = \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{|i - j| \left(\frac{N_i}{n} s_i + \frac{N_j}{n} s_j\right)}{N_i + N_j}$$

for $N_i \neq 0$ and $N_j \neq 0$

The Gray Level Size Zone Matrix

In the Gray Level Size Zone Matrix [3] calculations the voxels with ‘not a number’ value were excluded from the analysis. Let P_{ij} denotes the (i,j) entry in the Gray Level Size Zone Matrix, i - gray value, j - size, n_r - number of homogeneous areas inside a studied structure and $p_{ij} = P_{ij}/n_r$.

$$gray\ level\ non-uniformity = \frac{1}{n_r} \sum_{i=1}^{M} \left(\sum_{j=1}^{N} P_{ij}\right)^2$$

$$size\ zone\ non-uniformity = \frac{1}{n_r} \sum_{j=1}^{N} \left(\sum_{i=1}^{M} P_{ij}\right)^2$$

$$small\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij}}{j^2}$$

$$large\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij} \cdot j^2$$

$$low\ gray\ level\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{P_{ij}}{i^2}$$

$$high\ gray\ level\ zone\ emphasis = \frac{1}{n_r} \sum_{i=1}^{M} \sum_{j=1}^{N} P_{ij} \cdot i^2$$

$$small\ zone\ low\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2 \cdot j^2}$$

$$small\ zone\ high\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\ \cdot i^2}{j^2}$$

$$large\ zone\ low\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\ \cdot j^2}{i^2}$$

$$large\ zone\ high\ gray\ level\ emphasis = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}P_{ij} \cdot i^2 \cdot j^2$$

$$size\ precentage = \frac{n_r}{number\ of\ voxels\ in\ the\ studied\ structure}$$

Shape

To calculate shape features contours were transformed onto 1 mm isotropic grid. The volume and surface estimation was done using marching cubes algorithm implemented in the VTK library [4] (V - volume, A - surface).

$$compactness\ 1 = \frac{V}{(\pi^2 A)^{3/2}}$$

$$compactness\ 2 = 36\pi\frac{V^2}{A^3}$$

$$spherical\ disproportion = \frac{A}{4\pi R^2}$$

where: R is the radius of a sphere with the same volume as the tumor.

$$sphericity = \frac{(36\pi V^2)^{1/3}}{A}$$

$$surface\ to\ volume\ ratio = \frac{A}{V}$$

$median\ thickness$ - median of distances of each voxel in the region of interest to its surface, calculated using distance transform
$SD\ thickness$ - standard deviation of distances of each voxel in the region of interest to its surface
$maximum\ 3D\ diameter$ - the largest pairwise Euclidian distance between voxels of the region of interest
$fractal\ dimension$ - calculated using box contouring technique and fixed grid scans excluding the voxels with ‘not a number’ value (Figure 3S) [5].

$$fractal\ dimension = -\frac{\ln\big(N(r)\big)}{\ln(r)} - I$$

where: r - size of the contouring box, $N(r)$ - number of boxes of size r containing at least one voxel, which belongs the studied structure, I - intercept.

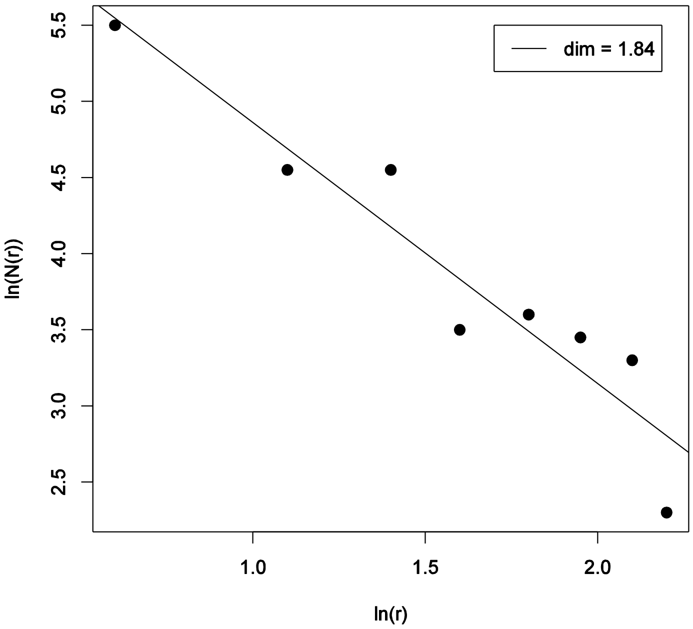


Figure 1S. Example of fractal dimension estimation

Wavelet

The wavelet transform was performed in 3D using Coiflet (‘coif1’) function from package PyWavelets (v. 0.2.2). The 5 histogram of intensities features and 29 texture features were calculated on 8 wavelet images, which resulted in 272 wavelet features.

Results

		Oropharyngeal carcinoma			
		Training		Validation	
Model	**Parameters of the model**	**CI**	**p-value**	**CI**	**p-value**
Radiomics	HHH large zone high gray-level emphasis LLL sum entropy LLH difference variance	0.79	< 0.001	0.87	< 0.001
Clinical	Stage Volume HPV	0.83		0.75	
Radiomics	HHH large zone high gray-level emphasis LLL sum entropy LLH difference variance	0.79	< 0.001	0.87	< 0.046
Radiomics + clinical	LLH difference variance Volume HPV	0.83		0.80	

Table 1S. The three local control prognostic models (radiomics, clinical and mixed) developed for the mixed cohort of patients and tested in oropharyngeal carcinoma. The model prognostic value is quantified by concordance index (CI). Wilcoxon test p-values are given for the comparison of the models performance.

Model	**Parameters of the model**	**HR**	**p-value**	**CI**
Radiomics	HHH large zone high gray-level	1.30	0.077	0.75
	emphasis LLL sum entropy	2.49	0.045	
	LLH difference variance	0.35	0.097	
Clinical	Stage	2.48	0.172	0.79
	Volume	1.02	0.021	
	HPV	0.10	0.002	
Radiomics + clinical	LLH difference variance	0.50	0.141	0.80
	Volume	1.02	0.007	
	HPV	0.11	0.004	
Clinical 2	Volume	1.02	< 0.001	0.77
	HPV	0.09	< 0.001	

Table 2S. Details of the tumor local recurrence prognostic models, HR – hazard ratio, p-value – from the multivariate Cox model, CI – concordance index.

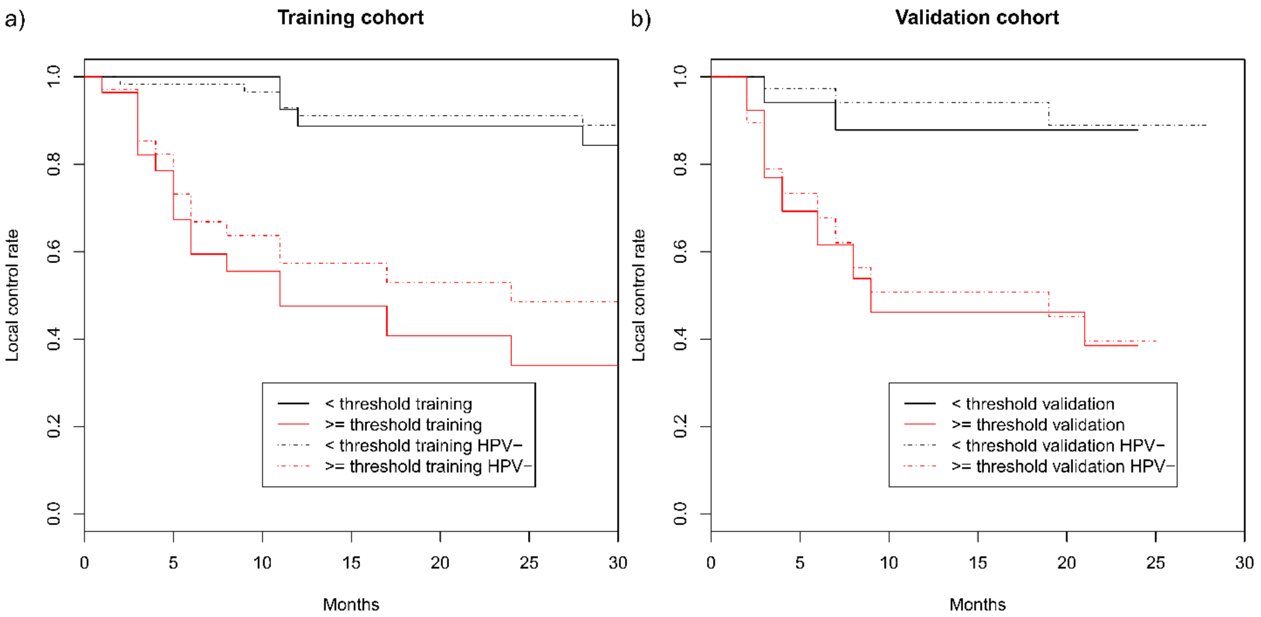


Figure 2S. Radiomics-based local tumor control model. Survival curves split significantly (G-rho test p-value < 0.001) in both training and validation cohorts as well as the subgroup of HPV negative tumors based on the optimal sensitivity-specificity threshold of the training cohort.

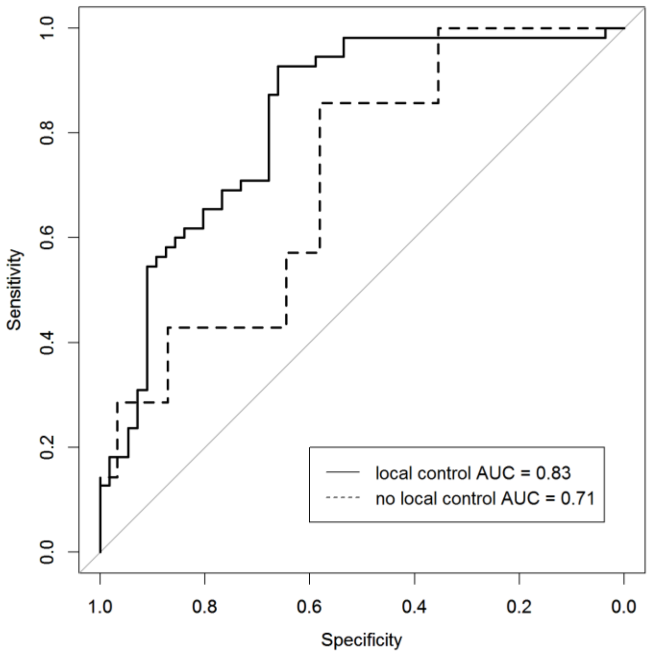


Figure 3S. The receiver operating characteristics for HPV status prediction using radiomics in patient cohort with and without local tumor control.

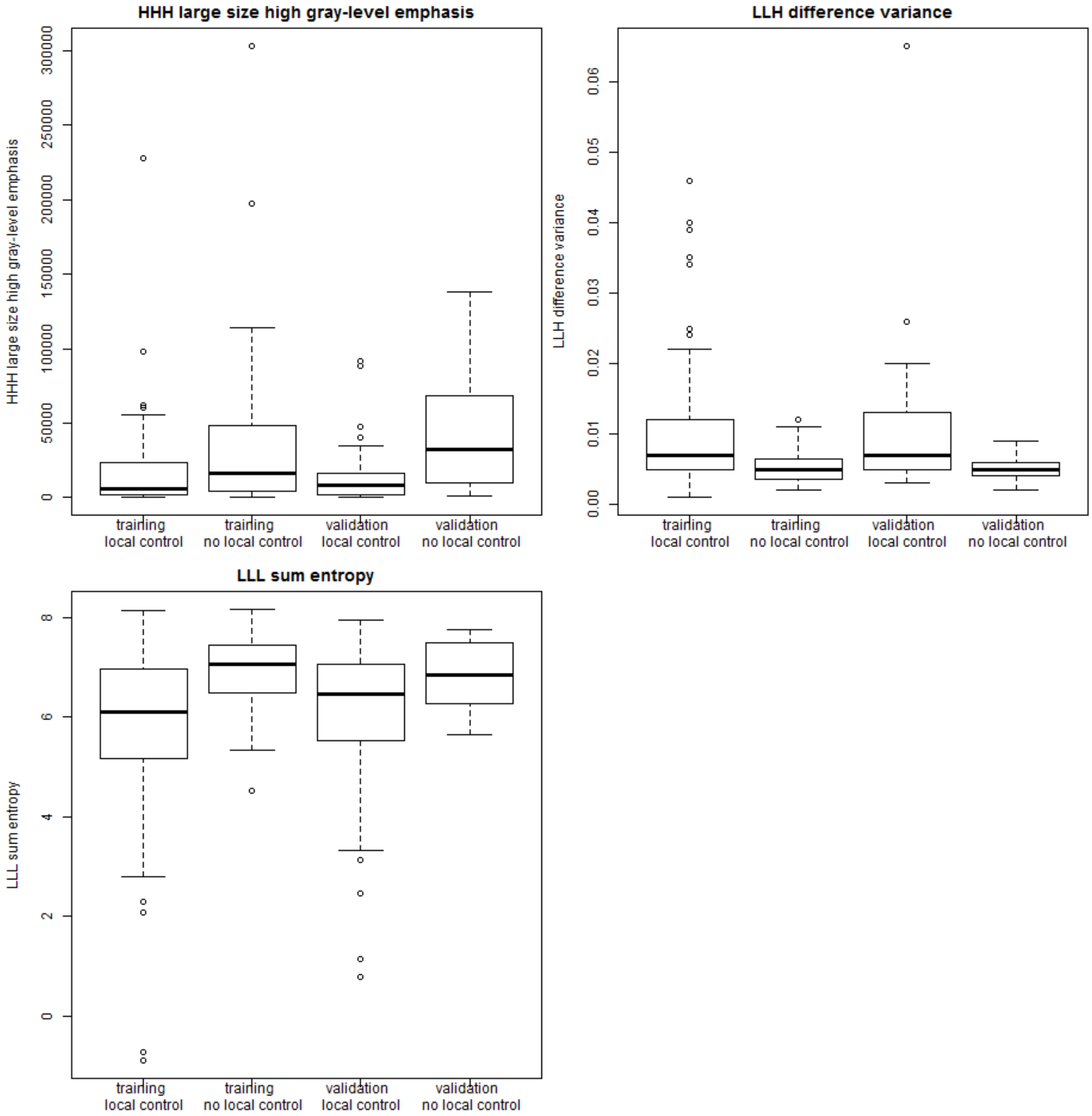


Figure 4S. Pre-treatment CT radiomic features prognostic for local tumor control

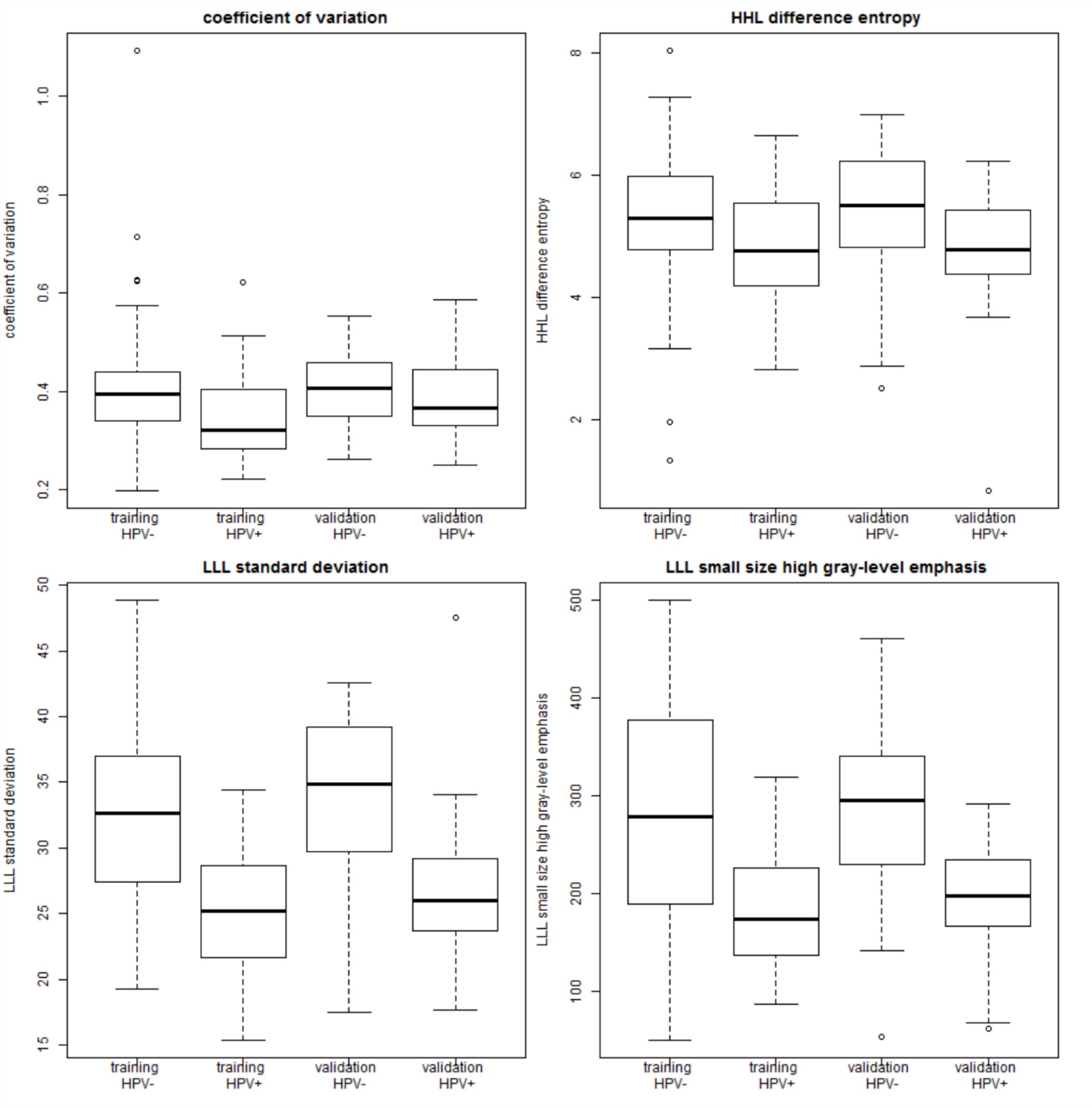


Figure 5S. CT radiomic features for prediction of status

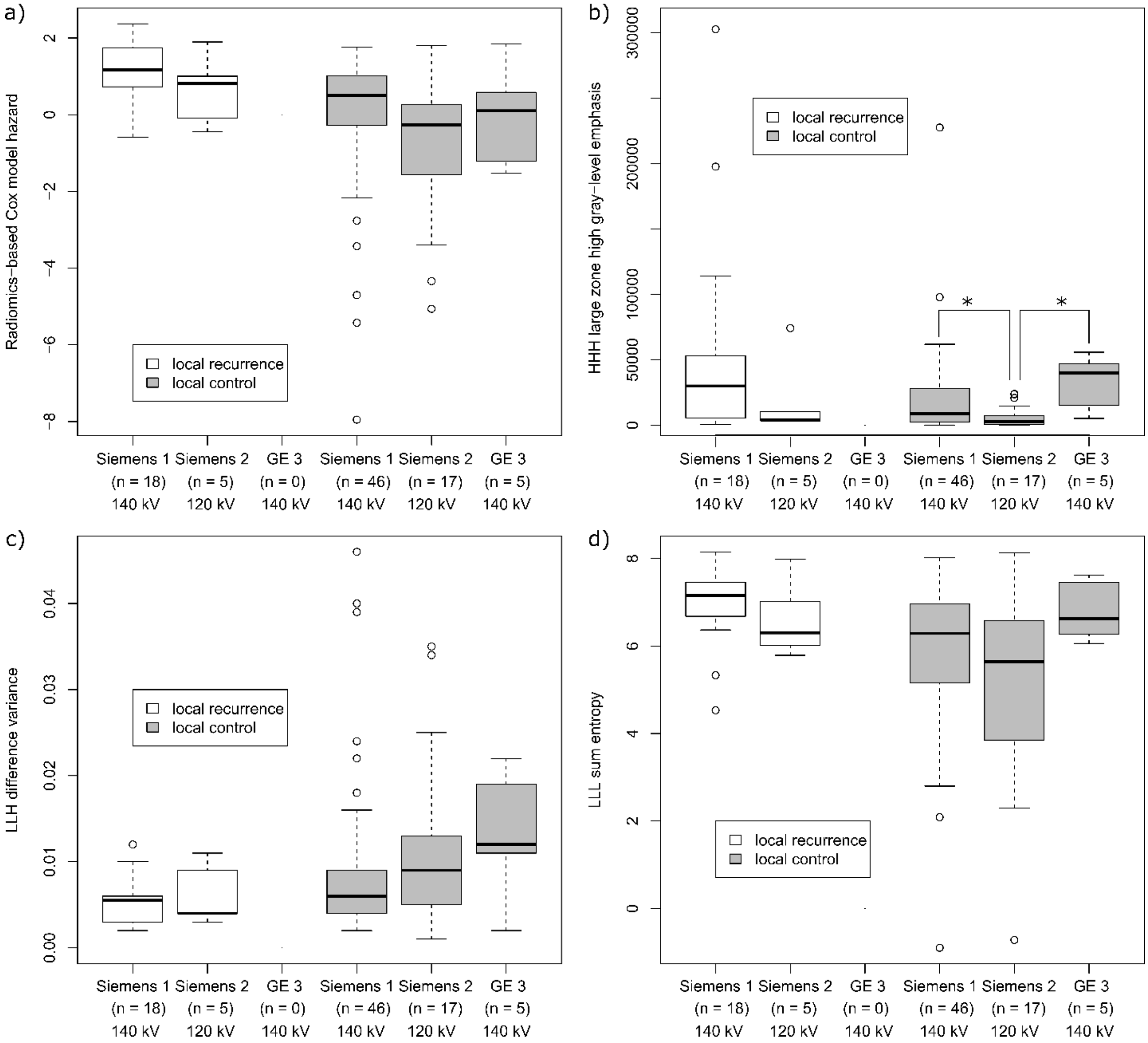


Figure 6S. The influence of CT scanner type on radiomics-based local tumor control model. Only the parameter HHH large zone high gray-level emphasis differs between the scanners in tumors where local control was achieved (Wilcoxon test p-value < 0.05).

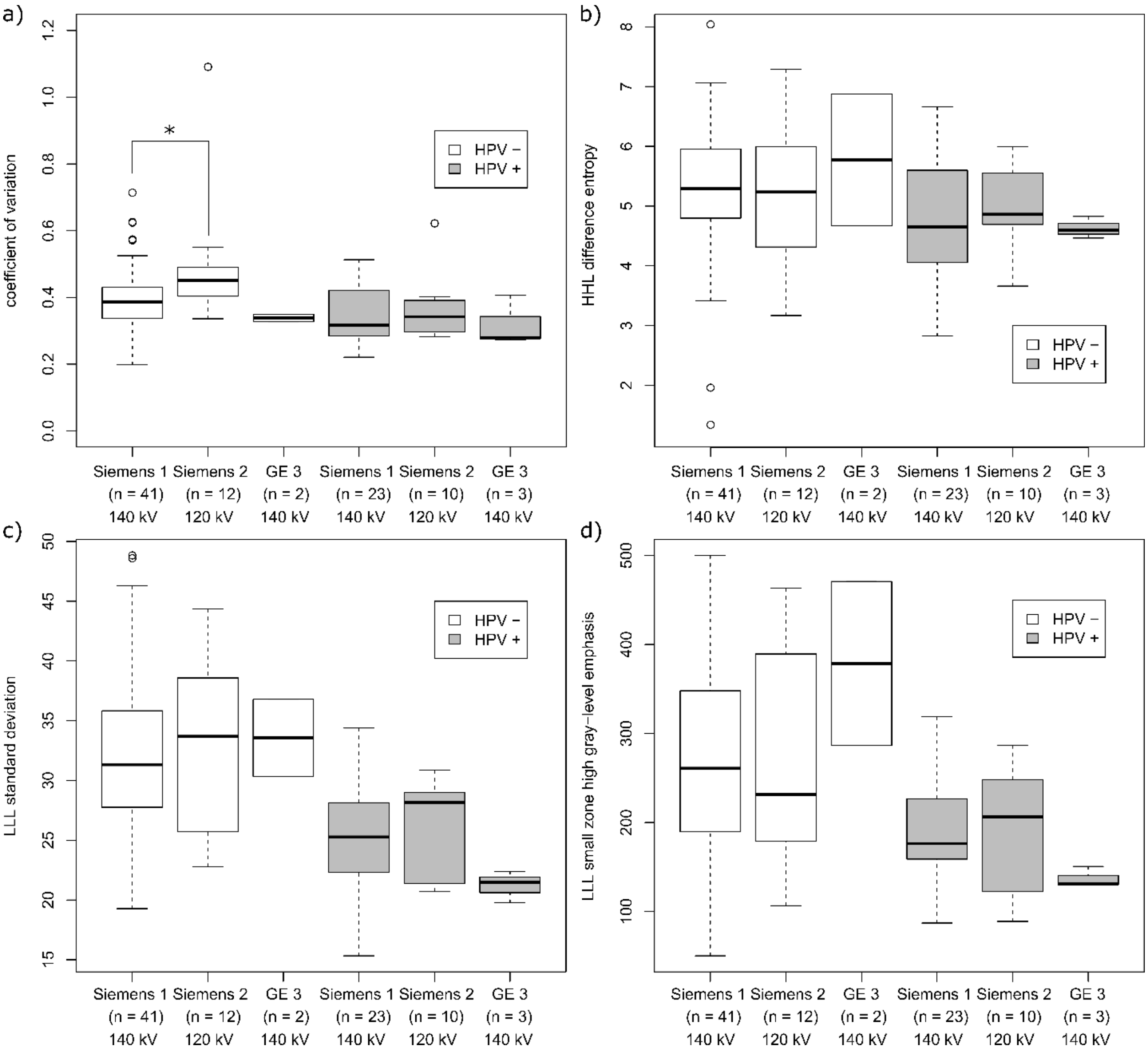


Figure 7S. The influence of CT scanner type on radiomics-based HPV prediction model. Only the parameter coefficient of variation showed a significant difference between the scanners in HPV negative tumors (Wilcoxon test p-value < 0.05).

References

[1] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classiification. IEEE Trans Syst Man Cybern 1973;3:610-621

[2] Amadasun M, King R. Texural Features Corresponding to Textural Properties. IEEE Trans Syst Man Cybern 1989;19(5)

[3] Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J, Mari JL. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. Pattern Recognition Inf Process 2009:140-145

[4] Schroeder W, Martin K, Lorensen B. The Visualization Toolkit (4th ed.), Kitware, 2006.

[5] Napolitano A, Ungania S, CannataV. Fractal Dimension Estimation Methods for Biomedical Images, in MATLAB - A Fundamental Tool for Scientific Computing and Engineering Applications. InTech, 2012, volume 3

3.6 Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma

Bogowicz Marta[1], Riesterer Oliver[1], Stark Louisa S[1], Studer Gabriela[1], Unkelbach Jan[1], Guckenberger Matthias[1], Tanadini-Lang Stephanie[1]

[1] Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich Switzerland.

Status of the manuscript:

published Acta Oncol. 2017, in press

Author contribution Marta Bogowicz:

- PET image segmentation
- Implementation of different techniques for features selection and classification
- Full data analysis: radiomics calculation, training of local tumor control prognostic models and comparison between those models (results for Figures 1-2)
- Preparation of all supplementary files
- Manuscript writing, revision, editing and proofreading.

ACTA ONCOLOGICA, 2017
https://doi.org/10.1080/0284186X.2017.1346382

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma

Marta Bogowicz, Oliver Riesterer, Luisa Sabrina Stark, Gabriela Studer, Jan Unkelbach,
Matthias Guckenberger and Stephanie Tanadini-Lang

Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Zurich, Switzerland

ABSTRACT

Purpose: An association between radiomic features extracted from CT and local tumor control in the head and neck squamous cell carcinoma (HNSCC) has been shown. This study investigated the value of pretreatment functional imaging (18F-FDG PET) radiomics for modeling of local tumor control.

Material and Methods: Data from HNSCC patients ($n = 121$) treated with definitive radiochemotherapy were used for model training. In total, 569 radiomic features were extracted from both contrast-enhanced CT and 18F-FDG PET images in the primary tumor region. CT, PET and combined PET/CT radiomic models to assess local tumor control were trained separately. Five feature selection and three classification methods were implemented. The performance of the models was quantified using concordance index (CI) in 5-fold cross validation in the training cohort. The best models, per image modality, were compared and verified in the independent validation cohort ($n = 51$). The difference in CI was investigated using bootstrapping. Additionally, the observed and radiomics-based estimated probabilities of local tumor control were compared between two risk groups.

Results: The feature selection using principal component analysis and the classification based on the multivariable Cox regression with backward selection of the variables resulted in the best models for all image modalities ($CI_{CT} = 0.72$, $CI_{PET} = 0.74$, $CI_{PET/CT} = 0.77$). Tumors more homogenous in CT density (decreased $GLSZM_{size\_zone\_entropy}$) and with a focused region of high FDG uptake (higher $GLSZM_{SZLGE}$) indicated better prognosis. No significant difference in the performance of the models in the validation cohort was observed ($CI_{CT} = 0.73$, $CI_{PET} = 0.71$, $CI_{PET/CT} = 0.73$). However, the CT radiomics-based model overestimated the probability of tumor control in the poor prognostic group (predicted = 68%, observed = 56%).

Conclusions: Both CT and PET radiomics showed equally good discriminative power for local tumor control modeling in HNSCC. However, CT-based predictions overestimated the local control rate in the poor prognostic validation cohort, and thus, we recommend to base the local control modeling on the 18F-FDG PET.

ARTICLE HISTORY
Received 1 May 2017
Accepted 14 June 2017

Introduction

Head and neck cancer squamous cell carcinomas (HNSCC) show a heterogeneous response to radiochemotherapy with locoregional control and five-year overall survival ranging from below 50% to 80% [1–4]. Few biomarkers related to tumor response to a therapy are already available for this tumor entity, for example: human papillomavirus infection [1–2], overexpression of epidermal growth factor receptor [5], tumor hypoxia or overexpression of cyclin D1 [6]. Apart from the interpatient variations, an intratumor heterogeneity in treatment response is observed. The tissue-based biomarkers from tumor biopsies may not account for intratumor heterogeneity, except for selected markers and marker profiles [7], and thus, other approaches should be investigated.

Radiation oncology is an imaging driven discipline incorporating precision medicine since many years. Each treatment plan is personalized based on pretreatment imaging. So far images collected in radiation oncology are a widely unused source of data, which could have a much higher impact on radiation treatment. However, an expert-based, observer specific analysis would limit the applicability of a decision support system. Radiomics is a method for quantitative analysis of medical images [7,8]. It provides tools for comprehensive image analysis on different scales: shape, intensity, texture and transform-based analysis. Its potential as a prognostic or predictive biomarker has been studied for different image modalities and tumor entities [9–11].

On the level of anatomical imaging, pretreatment CT radiomics was found to be prognostic for local tumor control [12] and overall survival [11,13–15] in head and neck cancer, significantly stratifying patients into low- and high-risk groups. Those studies showed that tumors, which are heterogeneous in terms of CT density are associated with a worse prognosis. The analysis of heterogeneity in functional

CONTACT Marta Bogowicz marta.nesteruk@usz.ch Department of Radiation Oncology, University Hospital Zurich, Rämistrasse 100, Zurich, 8091, Switzerland

Supplemental data for this article can be accessed here.

imaging could potentially further improve the imaging-based prognostic models. In head and neck cancer, texture of dynamic contrast-enhanced MRI (DCE-MRI) showed differences throughout the course of radiochemotherapy [16]. However, these differences were not correlated with local treatment failure. Another DCE-MRI study suggested that the histogram analysis provides parameters prognostic for progression-free survival and overall survival in stage IV HNSCC [17]. Conflicting results have been published regarding the role of pretreatment 18F-FDG PET as a prognostic factor in head and neck cancer. Some retrospective data showed an association between disease-free survival and increased maximum as well as mean standardized uptake value (SUV), whereas other groups did not observe such a correlation [18]. Recently, Vallieres et al. [19] tried to predict locoregional control using combined PET and CT radiomics. This study was not successful. Considering the increasing evidence of the association between CT radiomics and local tumor control [12], the mixed experiences using functional and anatomical imaging could be explained by other non-tumor related factors influencing nodal recurrence or too large number of studied radiomic features in relation to number of included patients.

In this work, we evaluate whether the metabolic information captured by 18F-FDG PET improves the prediction of local tumor control compared to the anatomical information from CT. Additionally, we investigate if the combination of both CT and PET can further improve the modeling.

Material and methods

Studied population and imaging protocol

Two cohorts of head and neck squamous cell carcinoma (HNSCC) patients were analyzed, a retrospective cohort of 121 patients (training cohort) with stage-III or stage-IV HNSCC and 51 patients (validation cohort) with the same tumor stages from a prospective phase-II study (NCT01435252). Inclusion criteria were as follows: treatment with definitive radiochemotherapy, contrast-enhanced planning CT and diagnostic 18F-FDG PET images available, no induction chemotherapy, no surgery (biopsy allowed), no previous or other simultaneous malignancies. Patients in the training cohort were treated with 70 Gy and concomitant weekly cisplatin (40 mg/m^2, up to 7 cycles) or cetuximab (loading dose 400 mg/m^2 followed by 250 mg/m^2 weekly), whereas the validation cohort received triple therapy consisting of radiotherapy (70 Gy) and weekly concurrent cisplatin/cetuximab (same doses as in training cohort) with or without consolidation cetuximab (500 mg/m^2 biweekly × 6). In both cohorts, the elective lymph nodes were irradiated with 54 Gy. The radiation dose was delivered using intensity modulated and simultaneous integrated boost technique in all patients. Follow-up included examinations by otorhinolaryngology specialists every 2–3 months and repetitive 18F-FDG PET/CTs. Local recurrence was proven by a biopsy in the region of primary tumor. Further details on the patient population are presented in Table 1.

All patients underwent a diagnostic 18F-FDG PET scan and a contrast-enhanced planning CT. On average, the scans were performed within 25 days. In the case of PET imaging, an activity of 178–513 MBq was administered intravenously 1 h prior to the scan and after the measurement of blood sugar level. In the retrospective cohort, 2D or 3D iterative image reconstruction was used, whereas the images of the validation cohort were reconstructed with 3D algorithm. Before the CT imaging, iodine contrast was injected intravenously. Its dose varied among the patients in the training cohort but was standardized in the validation cohort. Also the scanner type and the scanning parameters: tube current and kV, varied between the patients in the training cohort. Specific details as well as the information about image resolution are shown in Table 2.

Table 1. Detailed patients characteristic.

	Training cohort	Validation cohort
Number of patients	121	51
Number of recurrences	35 (29%)	15 (29%)
Median follow-up [months]	64	22
Age [years][a]	59 (34–73)	58 (47–75)
HPV status		
p16 positive	25 (21%)	28 (55%)
p16 negative	37 (31%)	23 (45%)
not known	59 (48%)	0 (0%)
Tumor stage		
T1/T2	45 (37%)	6 (12%)
T3/T4	76 (63%)	45 (88%)
Nodal stage		
N0	26 (22%)	9 (18%)
N1	10 (8%)	3 (6%)
N2	81 (67%)	37 (72%)
N3	4 (3%)	1 (2%)
Tumor site		
Oropharynx	80 (66%)	28 (55%)
Hypopharynx	24 (20%)	8 (16%)
Larynx	10 (8%)	7 (13%)
Oral cavity	7 (6%)	8 (16%)

[a] Minimum and maximum age.

Image preprocessing and radiomics

The primary tumor was separately segmented in the CT and PET images. The CT segmentation was performed manually. In all cases, two radiation oncologists, both having more than 10 years of experience, were involved in the process. Contours were later postprocessed for the presence of metal artifacts to exclude non-tumor related effects. If a certain tumor slice was affected by any artifacts, the entire tumor contour was erased from that slice. Tumors with more than 50% of volume not suitable for the analysis were not included in the study. Additionally, the voxels outside of soft tissue Hounsfield unit (HU) range (−20 HU to 180 HU) were discarded. The tumor in the PET image was autosegmented using a gradient-based method implemented in MIMVISTA (MIM Software Inc., Cleveland, OH). The contours and images were linearly interpolated to cubic voxels. The voxel size corresponded to the largest dimension in the original dataset: 3.3 mm in CT (slice thickness) and 5.5 mm in PET (in-plane resolution).

A fully DICOM compatible, in-house developed radiomics software implementation written in the Python

Table 2. The scanning parameters for training and validation cohorts and two image modalities.

Modality	Scanning parameters	Training cohort	Validation cohort
CT	Scanners	Siemens SOMATOM Volume zoom ($n = 83$) Siemens SOMATOM Definition AS ($n = 23$) GE Discovery STE ($n = 15$)	Siemens SOMATOM Definition AS ($n = 51$)
	Slice thickness [mm]	1.25–3.30	2
	In-plane resolution [mm]	0.85–1.95	0.98–1.56
	kV	120;140	120
	mAs	60–450	183–450
PET	Scanners	GE Discovery STE ($n = 51$) GE Discovery 690 ($n = 3$) GE Discovery HR ($n = 24$) GE Discovery LS ($n = 20$) GE Discovery RX ($n = 23$)	GE Discovery STE ($n = 35$) GE Discovery 690 ($n = 7$) GE Discovery RX ($n = 9$)
	Reconstruction	2D ($n = 61$) 3D ($n = 60$)	2D ($n = 0$) 3D ($n = 51$)
	Slice thickness [mm]	3.3; 4.3	3.3
	In-plane resolution [mm]	2.7–5.5	2.7–5.5
	Administered 18F-FDG activity [MBq]	351 (237–513)[a]	326 (178–406)[a]

[a] Minimum and maximum administered activity.

programing language was used. It provides 3D implementation of feature extraction methods for four types of features:

- shape ($n = 18$)
- intensity ($n = 19$)
- texture: the Gray Level Co-occurrence Matrix ($n = 26$), the Neighborhood Gray Tone Difference Matrix ($n = 4$) and the Gray Level Size Zone Matrix ($n = 14$)
- wavelet transform using Coiflet function (8 sub-bands, $n = 488$).

For each patient, 569 CT radiomic features and 569 PET radiomic features were extracted from the tumor volumes. For the texture and wavelet analysis, images were discretized to a fixed bin size of 5 HU and 0.25 SUV in the CT and PET, respectively. This choice of bin sizes resulted in similar numbers of bins in both modalities – on average 35 and 36 bins were analyzed in CT and PET, respectively. The details of the implementation are presented in the supplement.

Prognosis of local tumor control using radiomics

Three separate models to predict local tumor control were trained: CT based, PET based and the model comprising both CT and PET radiomic features. The model training consisted of two steps: feature selection and classification. Several feature selection and classification methods were used to reduce the model training method bias on the comparison between modalities and to train the best possible model. In the feature selection step, we considered:

- the principal component analysis (PCA) combined with univariable Cox regression,
- the Pearson correlation between the features and the principal components,
- the average clustering,
- the mutual information method,
- the minimum redundancy maximum relevance method.

The classification was performed using:

- the multivariable Cox regression with backward selection of the variables,
- the least absolute shrinkage and selection operator (LASSO),
- the random forest.

All the methods are described in the supplement. Per image modality, the model characterized by the largest concordance index (CI) in the 5-fold cross validation in the training cohort and the smallest number of included features was selected for further validation and comparison.

The risk group stratification was performed based on the trained models. The thresholds were chosen to obtain the same level of sensitivity and specificity in the receiver operating characteristic curve for local tumor control at 18 months.

Models validation and comparison

The radiomics-based local tumor control prognostic models, based on different imaging modalities, were verified and compared in the independent validation cohort. First, the differences in CI were investigated using Wilcoxon test (p-value < 0.05) and the bootstrap method with 100 randomly selected samples to calculate the CI distribution. Next, the risk group stratification was studied with the G-rho test. Finally, the linear dependence of the predicted and observed tumor control probabilities at 18 months for two risk groups was determined, to compare the stratifications from the different models. In this regard, a good model is defined by a linear dependence with slope equals 1.

Results

Radiomics-based models of local tumor control

Using different combinations of five feature selection and three classification methods, 15 local tumor control prognostic models were trained per image modality. They were compared based on the 5-fold cross-validated concordance index in the training cohort and the number of included features (Supplement Table 1S). The combination of feature selection using the PCA and univariable modeling and classification using the Cox regression with backward selection resulted in the least complicated model with the best discriminative power. This hold true for both PET and CT modeling.

The PCA resulted in the preselection of 12 CT features, five PET features and 11 features from the combined CT and PET feature set. Final CT- and PET-based tumor local control models consisted of two radiomic features (CT: $\text{GLSZM}_{\text{size\_zone\_entropy}}$, HLH $\text{intensity}_{\text{energy}}$; PET: spherical disproportion, $\text{GLSZM}_{\text{SZLGE}}$; see discussion paragraphs two and three for the interpretation of the features). The combined model comprised two features, a combination of CT and PET features, which were also selected in the single modality modeling (CT HLH $\text{intensity}_{\text{energy}}$ and PET $\text{GLSZM}_{\text{SZLGE}}$). All models showed a similar discriminative power in the training cohort using 5-fold cross-validation: $\text{CI}_{\text{CT}} = 0.72$, $\text{CI}_{\text{PET}} = 0.74$, $\text{CI}_{\text{PET/CT}} = 0.77$.

Validation of the local tumor control models and their comparison

The comparison of the local tumor control models was based on the data from the validation cohort. Studied radiomics-based models did not differ significantly in terms of discriminative power: $\text{CI}_{\text{CT}} = 0.73$, $\text{CI}_{\text{PET}} = 0.71$, $\text{CI}_{\text{PET/CT}} = 0.73$. All models significantly stratified patients into low- and high-risk groups of tumor recurrence (Figure 1). For the validation cohort, the use of combined PET and CT radiomics did not bring an added value to local tumor control modeling in HNSCC compared to single-modality radiomics.

There was a significant correlation ($r = 0.62$, p-value < 0.001) between the rankings of the patients in PET and CT radiomics models. However, the stratification of patients into low- and high-risk groups of tumor recurrence was different in 30% of the cases. Although this concerned mostly the patients close to the stratification thresholds, a group of seven patients, who were ranked low based on the CT model and high based on the PET model, was identified (Figure 2(a)). The observed and model-based estimated probabilities of local tumor control at 18 months were compared for both risk groups (Figure 2(b)). The PET-based model gave a better estimation with a linear regression slope of 1.12 in comparison to 0.58 in the CT-based model. The latter model overestimated the probability of tumor control in the poor prognosis group. Those results are consistent with the differential classification of the aforementioned seven patients.

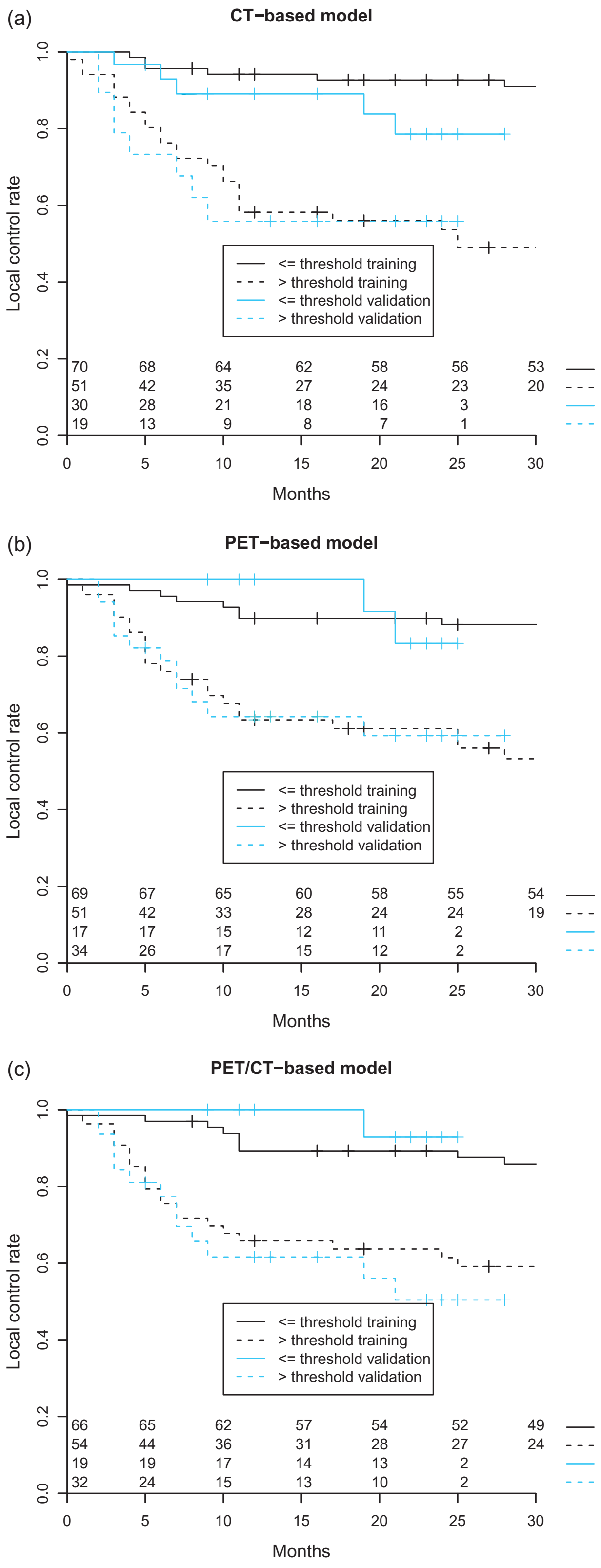


Figure 1. Radiomics-based local tumor control prognostic models: (a) CT-based radiomics, (b) PET-based radiomics, (c) CT- and PET-based radiomics. Tumor control curves split significantly (G-rho test p-value < 0.05) in both training and validation cohorts based on the optimal sensitivity-specificity thresholds at 18 months.

Discussion

Identification of new biomarkers in head and neck cancer is crucial to improve heterogeneous treatment outcome after definitive radiochemotherapy. A potential source of novel biomarkers could be radiomics, which can capture spatial tumor heterogeneity in terms of morphology or tumor biology, depending on the analyzed image type. This study showed that both heterogeneity in tumor CT density and in tumor metabolic activity, measured by 18F-FDG PET, are prognostic for local tumor control in HNSCC. CT radiomics

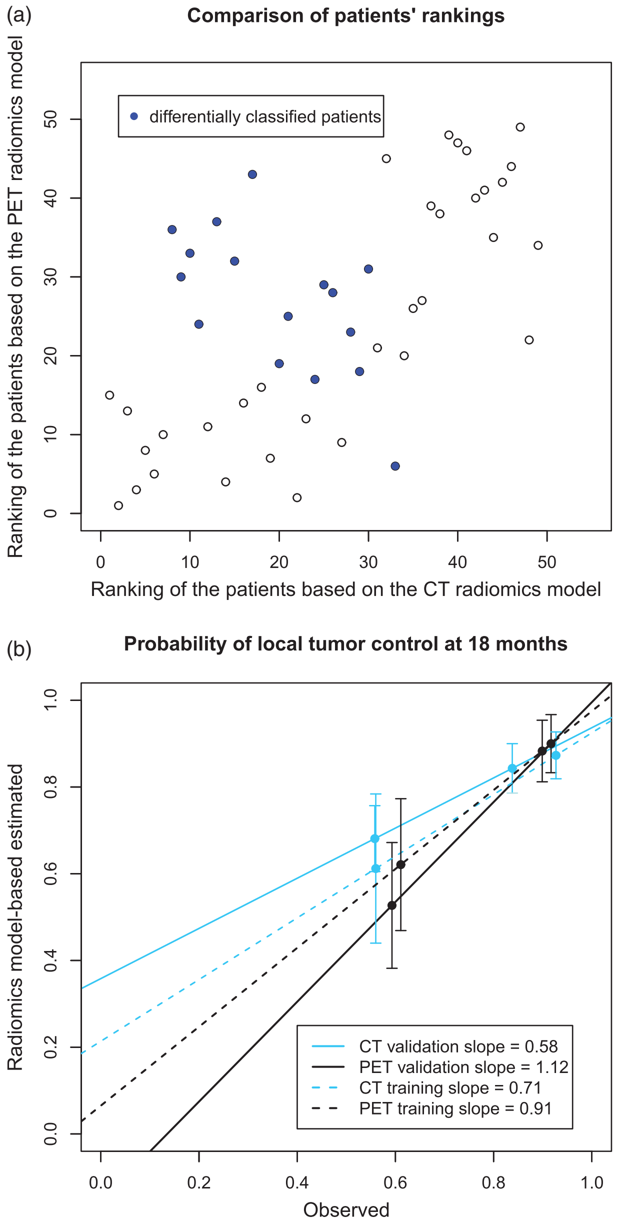


Figure 2. The comparison of risk group stratification in the validation cohort by CT- and PET-based radiomics models (a) the repeatability of patients' ranking based on CT and PET radiomics models and differences in risk group stratification, (b) the relation between observed and radiomics model-based estimated probabilities of local tumor control at 18 months for risk groups defined using the radiomics models.

and PET radiomics-based models showed a similar discriminative power (CI around 0.7) in the training as well as in the independent validation datasets. They also significantly stratified patients into low- and high-risk groups of tumor recurrence. However, the probability of tumor control in poor prognostic cohort was overestimated using the CT-based model, which was not observed in the PET-based radiomics model. Overall, the multimodality radiomics combining PET and CT did not improve the local tumor control modeling.

On the level of morphological imaging, head and neck tumors characterized by increased heterogeneity in CT density (increased $\mathrm{GLSZM}_{\mathrm{size\_zone\_entropy}}$) were associated with worse prognosis (Supplement Figure 2S). This result is consistent with previously published studies investigating local control [12] and overall survival [13,15] as endpoints. To understand the underlying biological processes of these findings, it would be of great interest to investigate if radiomic features corresponding to heterogeneity correlate with histopathological stainings of cell proliferation, microvessel density or hypoxia.

Regarding tumor metabolic activity, round tumors based on the PET signal autosegmentation (smaller spherical disproportion) with a focused region of high FDG uptake surrounded by a rim of low FDG uptake (higher $\mathrm{GLSZM}_{\mathrm{SZLGE}}$) were associated with a better prognosis (Supplement Figure 2S). This confined FDG uptake could correlate with a less infiltrative phenotype. On the other hand, a rim of low FDG uptake could correspond to inflammation, which has been previously correlated to favorable prognosis in HNSCC [20]. These radiomic features did not correlate with tumor volume, range of the SUV or maximum SUV. Additionally, in the studied cohorts of patients the standard PET measures, mean and maximum SUV, were not prognostic for tumor control. Moreover, other studies have shown that those standard PET measures correlate with tumor stage [18,21], whereas the predictions based on our PET radiomics model were independent of tumor stage (Supplement Figure 2S). This demonstrates, on a clinically relevant dataset, that radiomics is an added value in comparison to standard PET measures in terms of local tumor control prediction.

This study shows a link between CT radiomics, PET radiomics as well as combined PET and CT radiomics and local tumor control in head and neck cancer. It is contradictory to the study of Vallieres et al. [19] in which no correlation between locoregional control and PET/CT radiomics was found. To translate our results to locoregional control modeling, the inclusion of additional parameters, for example lymph node-specific radiomics, should be considered. In our study, the combination of PET and CT radiomics as well as single modality radiomics resulted in equally good discriminative power of the local tumor control.

The detailed comparison of the single modality models revealed that despite the significant stratification into risk groups by both modalities, the CT radiomics model substantially overestimates the probability of tumor control in the high-risk group. The subgroup of patients, for whom the risk-group classification depended on image modality, was identified. Most of those cases (70%) was affected by the CT artifacts in the tumor region. Although, the slices with artifacts were removed from the radiomics analysis this procedure seems to have influenced the results. Those slices could have contained a crucial information about tumor heterogeneity. Another source of error could be a variable contrast dose between training and validation cohorts. In such a setting, PET radiomics, despite its low resolution, appears to be better suited for the prognostic modeling in head and neck region.

Data in the training cohort were acquired using different scanners as well as different scanning and reconstruction settings, which might have influenced the modeling. Validation of the models in the independent dataset with standardized imaging protocol suggests the models' robustness, but it should be studied in more detail.

Further research in this area should be performed to address certain limitations of this study. A bigger dataset could be used for training a combined model to check for possible overfitting in the presence of the large number of input features. Additionally, this single institution study should be validated against multicenter data as well as inter-observer variability of gross tumor volume definition [22]. Moreover, it is known that not the entire tumor exhibits increased metabolic activity. Thus, parts of the tumor may have been missed by the autosegmentation. The PET-radiomics based on the treatment planning GTV could further improve the predictions. It would also eliminate a potential bias introduced by tumor progression between CT and PET scans [23]. Previous works [12–13] have shown that the inclusion of clinical parameters (tumor stage or volume) may result in more accurate predictions. However, so far other important clinical factors, such as smoking status or performance status, were not studied in combination with radiomics.

In conclusion, both CT and PET radiomics show potential to be a prognostic biomarker in head and neck cancer, with equally good discriminative power. However, the presence of CT artifacts may alter the CT-based predictions and therefore, radiomics-based outcome modeling in HNSCC appears to be more robust using 18F-FDG PET. Additionally, no benefit from combining CT and FDG-PET radiomics into one prognostic model was observed. Further validation of these results in a multicenter setting and against different CT artifacts correction protocols is needed.

Acknowledgments

The project was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant of the Matching Fund of the University of Zurich. The clinical study used as validation dataset was supported by a research grant from Merck (Schweiz) AG.

Disclosure of interest

No potential conflict of interest was reported by the authors.

Funding

The project was supported by the Clinical Research Priority Program Tumor Oxygenation of the University of Zurich, by a grant of the Matching Fund of the University of Zurich. The clinical study used as validation dataset was supported by a research grant from Merck (Schweiz) AG.

ORCID

Marta Bogowicz http://orcid.org/0000-0002-4747-5375

References

[1] Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med. 2010;363:24–35.

[2] Lassen P, Primdahl H, Johansen J, et al. Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol. 2014;113:310–316.

[3] Svahn MF, Munk C, Nielsen TS, et al. Trends in all-cause five-year mortality after head and neck cancers diagnosed over a period of 33 years. Focus on estimated degree of association with human papillomavirus. Acta Oncol. 2016;55:1084–1090.

[4] Bentzen J, Toustrup K, Eriksen JG, et al. Locally advanced head and neck cancer treated with accelerated radiotherapy, the hypoxic modifier nimorazole and weekly cisplatin. Results from the DAHANCA 18 phase II study. Acta Oncol. 2015;54:1001–1007.

[5] Riesterer O, Milas L, Ang KK. Use of molecular biomarkers for predicting the response to radiotherapy with or without chemotherapy. J Clin Oncol. 2007;25:4075–4083.

[6] Kang H, Kiess A, Chung CH. Emerging biomarkers in head and neck cancer in the era of genomics. Nat Rev Clin Oncol. 2015;12:11–26.

[7] Toustrup K, Sorensen BS, Metwally MA, et al. Validation of a 15-gene hypoxia classifier in head and neck cancer for prospective use in clinical trials. Acta Oncol. 2016;55:1091–1098.

[8] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–446.

[9] Hatt M, Tixier F, Pierce L, et al. Characterization of PET/CT images using texture analysis: the past, the present any future? Eur J Nucl Med Mol Imaging. 2017;44:151–165.

[10] Yip SSF, Aerts HJWL. Applications and limitations of radiomics. Phys Med Biol. 2016;61:150–166.

[11] Wong AJ, Kanwar A, Mohamed AS, et al. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res. 2016;5:371–382.

[12] Bogowicz M, Riesterer O, Ikenberg K, et al. CT radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. Int J Rad Onc Biol Phys. Forthcoming 2017.

[13] Aerts HJWL, Velazquez ER, Leijenaar RTH, etet al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

[14] Parmar C, Leijenaar RTH, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. Sci Rep. 2015;5:11004.

[15] Leijenaar RTH, Carvalho S, Hoebers FJ, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. Acta Oncol. 2015;54:1423–1429.

[16] Jansen JF, Lu Y, Gupta G, et al. Texture analysis on parametric maps derived from dynamic contrast-enhanced magnetic resonance imaging in head and neck cancer. WJR. 2016;8:90–97.

[17] Shukla-Dave A, Lee NY, Jansen JF, et al. Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases. Int J Radiat Oncol Biol Phys. 2012;82:1837–1844.

[18] Paidpally V, Chirindel A, Lam S, et al. FDG-PET/CT imaging biomarkers in head and neck squamous cell carcinoma. Imaging Med. 2012;4:633–647.

[19] Vallières M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. arXiv:1703.08516.

[20] De Schutter H, Landuyt W, Verbeken E, et al. The prognostic value of the hypoxia markers CA IX and GLUT 1 and the cytokines VEGF and IL 6 in head and neck squamous cell carcinoma treated by radiotherapy +/- chemotherapy. BMC Cancer. 2005;5:42.

[21] Nesteruk M, Lang S, Veit-Haibach P, et al. Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma. Radiother Oncol. 2015;117:125–131.

[22] Leijenaar RT, Carvalho S, Velazquez ER, et al. Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52:1391–1397.

[23] Jensen AR, Nellemann HM, Overgaard J. Tumor progression in waiting time for radiotherapy in head and neck cancer. Radiother Oncol. 2007;84:5–10.

Supplement

Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma

Marta Bogowicz[a], Oliver Riesterer[a], Luisa Sabrina Stark[a], Gabriela Studer[a], Jan Unkelbach[a], Matthias Guckenberger[a], Stephanie Tanadini-Lang[a]

[a]*Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Switzerland*

Radiomic features

The following approaches were used to determine radiomics: histogram of intensities analysis, texture analysis (the Gray Level Co-occurrence Matrix, the Neighborhood Gray Tone Difference Matrix, the Gray Level Size Zone Matrix and the Gray Level Run Length Matrix), shape analysis and wavelet transform analysis.

Histogram of intensities

The parameters from the histogram of intensities were calculated in the perfusion maps before a discretization. Let X denotes the intensities of the 3D image with N voxels. $\bar{X}$ - mean of X, N_g - number of gray levels in the image, p_i - the occurrence probability of gray level i.

1) $mean = \frac{1}{N}\sum_{i=1}^{N} X_i$
2) $standard\ devation = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2}$
3) $variance = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$
4) $coefficient\ of\ variation = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2}}{\bar{X}}$
5) $skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2}\right)^3}$
6) $kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})^2}\right)^4} - 3$
7) $mean\ absolut\ deviation = \frac{1}{N}\sum_{i=1}^{N}|X_i - \bar{X}|$
8) $robust\ mean\ absolut\ deviation = \frac{1}{N}\sum_{i=1}^{N_{10-90}}\left|X_{10-90,i} - \bar{X}_{10-90}\right|$

where: N_{10-90}- number of voxel in the range from and 10th percentile and 90th percentile, $\bar{X}_{10-90}$ - mean value of voxel in the range from 10th percentile and 90th percentile

9) $energy = \sum_{i=1}^{N} X_i^2$
10) $entropy = -\sum_{i=1}^{N_g} p_i \log_2 p_i$
11) $root\ mean\ square = \sqrt{\frac{energy}{N}}$
12) $uniformity = \sum_{i=1}^{N_g} p_i$
13) $minimum\ value = \min(X)$
14) $maximum\ value = \max(X)$
15) $median -$ the median value of X
16) $10^{th} percentile$ – 10th percentile of X
17) $90^{th} percentile$ – 90th percentile of X
18) interqurtile range = $90^{th} percentile - 10^{th} percentile$
19) $range = \max(X) - \min(X)$

The Gray Level Co-occurrence Matrix (GLCM)

The parameters from the Gray Level Co-occurrence Matrix [1] were calculated in all 26 directions with a distance of one voxel. The final parameters were the average of all directions. If one of the voxels had a ‘not a number’ value the pair was not taken into account in the calculations. Let P_{ij} denotes the (i,j) entry in the Gray Level Co-occurrence Matrix, N_g - number of gray tones in a studied structure, $P_{xi} = \sum_{j=1}^{N_g} P_{ij}$, $P_{yj} = \sum_{i=1}^{N_g} P_{ij}$, $P_{x+y}(k) = \sum_{j=1}^{N_g}\sum_{i=1}^{N_g} P_{ij}$, where $k = i + j$, $P_{x-y}(k) = \sum_{j=1}^{N_g}\sum_{i=1}^{N_g} P_{ij}$, where $k = |i - j|$.

20) $energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}^2$

21) $contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-j)^2 P_{ij}$

22) $correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP_{ij} - \mu_x\mu_y}{\sigma_x\sigma_y}$

where: $\mu_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} iP_{ij}$, $\mu_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} jP_{ij}$, $\sigma_x = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}(i-\mu_x)^2$,

$\sigma_y = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\left(j-\mu_y\right)^2$

23) $sum\ of\ squares = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-\mu)^2 P_{ij}$

where μ – mean of P

24) $inverese\ difference\ moment\ (homogeneity) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P_{ij}}{1+(i-j)^2}$

25) $inverese\ difference\ moment\ normalized = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P_{ij}}{1+\frac{(i-j)^2}{N_g^2}}$

26) $inverese\ difference\ = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P_{ij}}{1+|i-j|}$

27) $inverese\ difference\ normalized = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{P_{ij}}{1+\frac{|i-j|}{N_g}}$

28) $sum\ average = \sum_{i=2}^{2N_g} i \cdot P_{x+y}(i)$

29) $sum\ entropy = -\sum_{i=2}^{2N_g} P_{x+y}(i) \log_2 P_{x+y}(i)$

30) $sum\ variance = \sum_{i=2}^{2N_g} P_{x+y}(i) \cdot (i - sum\ average)^2$

31) $entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij} \log_2 P_{ij}$

32) $difference\ entropy = -\sum_{i=0}^{N_g-1} P_{x-y}(i) \log_2 P_{x-y}(i)$

33) $information\ measure\ of\ correlation\ 1\ (IMC1) =$

$$\frac{-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij} - \left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right)}{max\left\{\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{xi}\log_2 P_{xi}\right),\left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{yj}\log_2 P_{yj}\right)\right\}}$$

34) $information\ measure\ of\ correlation\ 2\ (ICM2) =$

$$\sqrt{1 - \exp\left[-2 - \left(-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{xi}P_{yj}\right) + \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_{ij}\log_2 P_{ij}\right]}$$

35) $maximal\ correlation\ coefficient\ (MCC) = \sqrt{\text{second largest eigenvalue of } \sum_{k=1}^{N_g} \frac{P_{ik}P_{jk}}{P_{xi}P_{ji}}}$

36) $joint\ maximum = \max(P_{ij})$

37) $joint\ average = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} iP_{ij}$

38) $difference\ average = \sum_{i=0}^{N_g-1} iP_{x-y}(i)$

39) $difference\ variance = \sum_{i=0}^{N_g-1} P_{x-y}(i) \cdot (i - difference\ average)^2$

40) $dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j|\, P_{ij}$

41) $inverese\ variance = 2 \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{(i-j)^2}$

42) $autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ijP_{ij}$

43) $cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^2 P_{ij}$

44) $cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^3 P_{ij}$

45) $cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^4 P_{ij}$

The Neighborhood Gray Tone Difference Matrix (NGTDM)

The Neighborhood Gray Tone Difference Matrix [2] was calculated based on 26 adjacent voxels. The voxels with 'not a number' value were excluded from the average over the neighborhood region. Let s_i denotes the *ith* entry in the Neighborhood Gray Tone Difference Matrix, N_i - the number of voxels having gray tone i, G - number of gray tones in a studied structure, n - number of studied voxels.

46) $coarseness = \left[\epsilon + \sum_{i=1}^{G} \frac{N_i}{n} s_i\right]^{-1}$

where ϵ is a small number to prevent coarseness becoming infinite.

47) $contrast = \left[\frac{1}{N_g(N_g-1)} \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{N_i}{n} \frac{N_j}{n} (i-j)^2\right] \left[\frac{1}{n} \sum_{i=1}^{G} s_i\right]$

where N_g is the total number of different gray levels present in the image.

48) $busyness = \frac{\sum_{i=1}^{G} \frac{N_i}{n} s_i}{\sum_{i=1}^{G} \sum_{j=1}^{G} i\frac{N_i}{n} - j\frac{N_j}{n}}$

for $\frac{N_i}{n} \neq 0$ and $\frac{N_j}{n} \neq 0$

49) $complexity = \sum_{i=1}^{G} \sum_{j=1}^{G} \frac{|i-j|\left(\frac{N_i}{n} s_i + \frac{N_j}{n} s_j\right)}{N_i + N_j}$

for $N_i \neq 0$ and $N_j \neq 0$

The Gray Level Size Zone Matrix (GLSZM)

In the Gray Level Size Zone Matrix [3] calculation the voxels with 'not a number' value were excluded from the analysis. Let P_{ij} denotes the (i,j) entry in the Gray Level Size Zone Matrix, i -

gray value, j - size, n_r - number of homogeneous areas inside a studied structure and $p_{ij} = P_{ij}/n_r$, $\mu_i = \sum_{i=1}^{M}\sum_{j=1}^{N} iP_{ij}$, $\mu_j = \sum_{i=1}^{M}\sum_{j=1}^{N} iP_{ij}$.

50) $gray\ level\ non-uniformity\ (GLNU) = \frac{1}{n_r}\sum_{i=1}^{M}\left(\sum_{j=1}^{N} P_{ij}\right)^2$

51) $size\ zone\ non-uniformity\ (SZNU) = \frac{1}{n_r}\sum_{j=1}^{N}\left(\sum_{i=1}^{M} P_{ij}\right)^2$

52) $small\ zone\ emphasis\ (SZE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{j^2}$

53) $large\ zone\ emphasis\ (LZE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot j^2$

54) $low\ gray\ level\ zone\ emphasis\ (LGZE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2}$

55) $high\ gray\ level\ zone\ emphasis\ (HGZE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2$

56) $small\ zone\ low\ gray\ level\ emphasis\ (SZLGE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}}{i^2\cdot j^2}$

57) $small\ zone\ high\ gray\ level\ emphasis\ (SZHGE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot i^2}{j^2}$

58) $large\ zone\ low\ gray\ level\ emphasis\ (LZLGE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{P_{ij}\cdot j^2}{i^2}$

59) $large\ zone\ high\ gray\ level\ emphasis\ (LZHGE) = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\cdot i^2\cdot j^2$

60) $size\ precentage = \frac{n_r}{number\ of\ voxels\ in\ the\ studied\ structure}$

61) $gray\ level\ variance = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}(i-\mu_i)^2 P_{ij}$

62) $size\ zone\ variance = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(j-\mu_j\right)^2 P_{ij}$

63) $size\ zone\ entropy = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij}\log(P_{ij})$

Shape

In the USZ implementation: to calculate shape features contours were transformed onto 1 mm isotropic grid. The volume and surface estimation was done using marching cubes algorithm implemented in the VTK library [4].

64) V – volume

65) A – surface

66) $compactness\ 1 = \frac{V}{(\pi^2 A)^{\frac{3}{2}}}$

67) $compactness\ 2 = 36\pi\frac{V^2}{A^3}$

68) $spherical\ disproportion = \frac{A}{4\pi R^2}$

where: R is the radius of a sphere with the same volume as the tumor.

69) $sphericity = \frac{(36\pi V^2)^{\frac{1}{3}}}{A}$

70) $asphericty = \left(\frac{1}{36\pi}\frac{A^3}{V^2}\right)^{\frac{1}{3}} - 1$

71) $surface\ to\ volume\ ratio = \frac{A}{V}$

72) $median\ thickness$ - median of distances of each voxel in the region of interest to its surface, calculated using distance transform
73) $SD\ thickness$ - standard deviation of distances of each voxel in the region of interest to its surface
74) $maximum\ 3D\ diameter$ - the largest pairwise Euclidian distance between voxels of the region of interest
75) $center\ of\ mass\ shift$ = geometric center of the mass – gray levels weighted center of the mass
76) $fractal\ dimension$ - calculated using box contouring technique and fixed grid scans excluding the voxels with 'not a number' value (Figure 1S) [5].

$$fractal\ dimension = -\frac{\ln(N(r))}{\ln(r)} - I$$

where: r - size of the contouring box, $N(r)$ - number of boxes of size r containing at least one voxel, which belongs the studied structure, I - intercept.

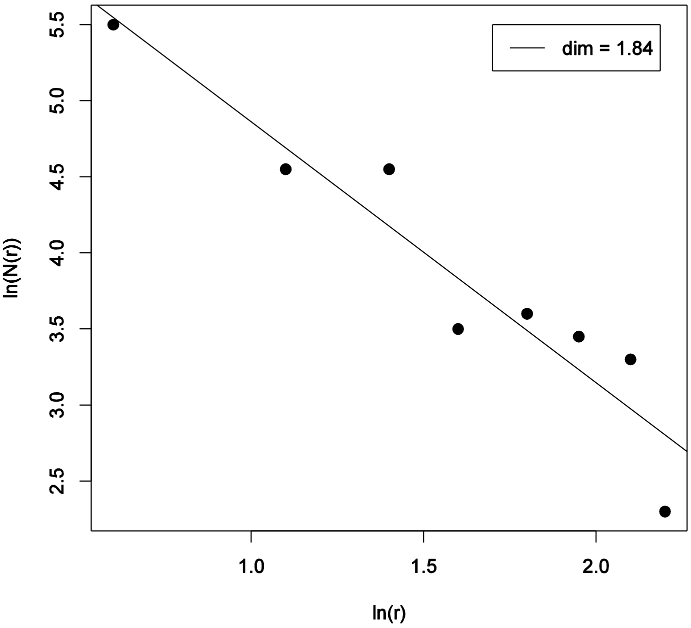


Figure 1S. Example of fractal dimension estimation

77) $major\ axis$ – the largest eigenvalue in the principal component analysis
78) $minor\ axis$ – the second largest eigenvalue in the principal component analysis
79) $least\ axis$ – the smallest eigenvalue in the principal component analysis
80) $alongation = \frac{minor\ axis}{major\ axis}$
81) $flatness = \frac{least\ axis}{major\ axis}$

Wavelet

The wavelet analysis was preformed using the Python library PyWavelets version 0.3.0 and the 'Coif1' wavelet as the filter. To account for the boundary effects the entire image set was filtered after the resizing to cubic voxels but before any image segmentation/resegmentation.

Feature selection

a) The principal component analysis combined with univariable Cox regression

The Horn method [6] was used to define the number of retained principal components. Next each of the radiomic features was assigned to the principal component based on its largest contribution to certain component [7]. Per principal component related group, the features with the highest concordance index (CI) and corresponding false discovery rate < 0.2 in the univariable Cox regression was selected.

b) The Pearson correlation between the features and the principal components

The Horn method [6] was used to define the number of retained principal components. Next, for each principal component one feature was selected to represent it. To that end we determined the feature that correlated the most (the largest Pearson correlation coefficient) with the principal component. In the contrary to the method a) no information about prognostic power of radiomic features was used.

c) The average clustering

The average clustering [8] was performed based on the Pearson correlations between radiomic features. The distance cut-off was set to 0.5. For each of the defined groups, the feature with the highest variance was selected to represent it.

d) *The mutual information*

The mutual information [9] between the radiomic features and tumor control probability was computed. The feature's selection threshold was defined as 80% of maximum mutual information between tumor control probability and any of the radiomic features.

e) The minimum redundancy maximum relevance

The ensemble maximum relevance minimum redundancy (MRMR) method was used [10]. It was partiality combined with the principal component analysis. The feature count was defined as the number of principal components, which explains 95% of data variance. The MRMR was repeated 1000 randomly selecting samples using bootstrap procedure. The redundancy between the features was defined as the Spearman correlation. Features, which achieved at least 80% selection rate were included in the final set.

Feature classification

a) *The multivariable Cox regression with backward selection of the variables*

The preselected features were used in the multivariable Cox regression with backward selection of the variables based on the Akaike information criterion [11]. Features, which had at least 80% selection rate in 5-fold cross-validation were chosen to the final model.

b) *The least absolute shrinkage and selection operator*

The least absolute shrinkage and selection operator [12] (100 times 5-fold cross-validated) was used for variable selection in multivariable Cox regression. Features, which had at least 70% selection rate were chosen to the final model.

c) *The random forest*

The random forest model was trained using 1000 trees and Breiman-Cutler permutation was calculate variable importance [13]. Features were selected based on the minimization of the out-of-bag error. The additional constrain was set for the maximal number of features n=5.

Results

a)

		Cox backward	**LASSO**	**RF**
PCA univariable Cox	**CI**	**0.72 (0.64-0.90)**	0.69 (0.59-0.85)	0.65 (0.54-0.88)
	n	2	1	3
PCA correlation	**CI**	0.68 (0.53-0.83)	0.69 (0.56-0.85)	0.55 (0.44-0.77)
	n	5	1	3
average clustering	**CI**	-	0.67 (0.47-0.90)	0.69 (0.64-0.80)
	n	-	1	5
mutual information	**CI**	0.71* (0.60-0.90)	0.71* (0.60-0.90)	0.68** (0.63-0.71)
	n	1	1	2
minimum redundancy maximum relevance	**CI**	0.71* (0.60-0.90)	0.71* (0.60-0.90)	0.68** (0.63-0.71)
	n	1	1	2

b)

		Cox backward	**LASSO**	**RF**
PCA univariable Cox	**CI**	**0.74** (0.62-0.80)**	0.70 (0.56-0.80)	0.70 (0.57-0.81)
	n	2	2	2
PCA correlation	**CI**	0.66 (0.54-0.74)	0.62 (0.55-0.71)	0.64 (0.60-0.74)
	n	3	1	3
average clustering	**CI**	0-67 (0.64-0.76)	0.67 (0.62-0.76)	0.54 (0.39-0.68)
	n	2	2	5
mutual information	**CI**	0.74** (0.62-0.80)	0.70 (0.66-0.74)	0.68 (0.57-0.80)
	n	2	1	3
minimum redundancy maximum relevance	**CI**	0.71 (0.57-0.81)	0.66 (0.62-0.79)	0.60 (0.50-0.66)
	n	1	1	3

Table 1S. The concordance indexes (CI) for radiomics-based local tumor control models trained with a different comibantions of features selection and classifications a) CT radiomics, b) PET radiomics. The results were based on the 5-fold cross-valdiation in the training cohort. The maxiumim and minimum CI obtianed in the cross-valdation are presented in the brackets. Additionaly, the number of features (n) included in the multivariable model is shown. The combination of the principal component analysis combined with univariable Cox regression and the multivarible Cox regression with backward selection of the variables resulted in the best models considering both CT and PET features (in bold). The combination of avarage clustering and Cox multivariable regression failed to optimize a model due to too large number of input variables (n=66). * the same final model containing 1 feature, ** the same final model containing 2 features. PCA univariable Cox - the principal component analysis combined with univariable Cox regression, PCA correlation - the Pearson correlation between the features and the principal components, Cox backward - the multivarible Cox regression with backward selection of the variables, LASSO - the least absolute shrinkage and selection operator, RF - the random forest.

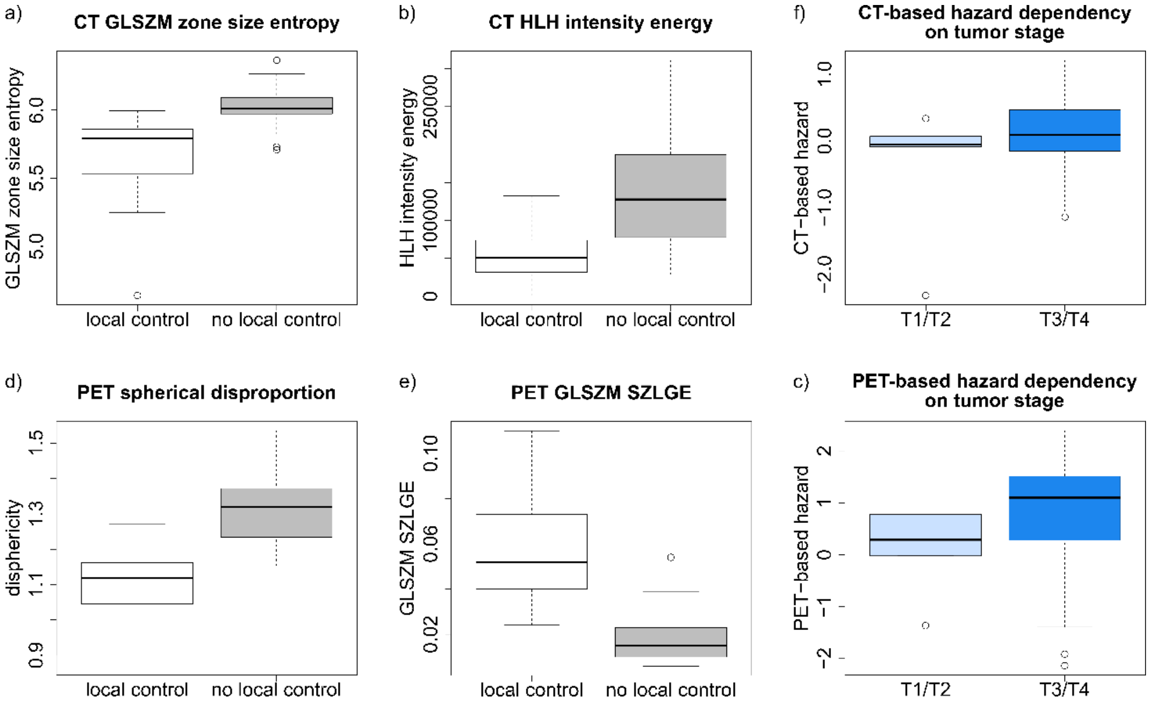


Figure 2S. The CT (a and b) and PET (d and e) radiomics signature prognostic for local tumor recurrence in head and neck cancer. All radiomic features differed significantly (Wilcoxon test p-value < 0.05) between the controlled tumors and recurrences. The predictions based on the radiomics model was independent from the tumor stage (c and f).

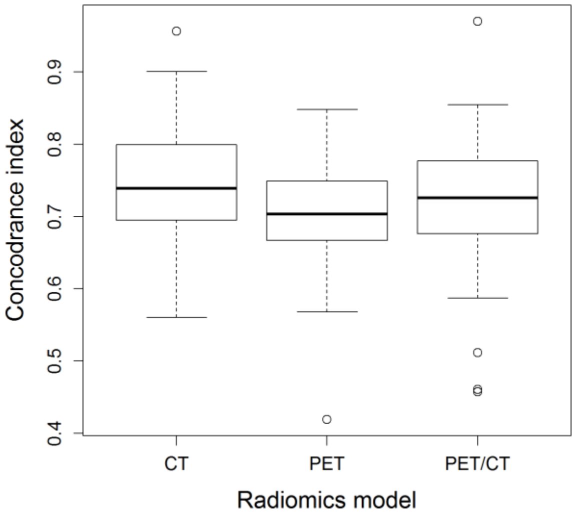


Figure 3. The concordance index distributions in the validation dataset for different local tumor control radiomics models. The distributions were obtained using the bootstrap method. None of the models showed a superior discriminative power.

References

[1] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;3:610-621.

[2] Amadasun M, King R. Texural features corresponding to textural properties. IEEE Trans Syst Man Cybern. 1989;19(5).

[3] Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J, Mari JL. Texture indexes and Gray Level Size Zone matrix application to cell nuclei classification. Pattern Recognition Inf Process. 2009;140-145.

[4] Schroeder W, Martin K, Lorensen B. The visualization toolkit (4th ed.). Kitware. 2006.

[5] Napolitano A, Ungania S, Cannata V, Fractal dimension estimation methods for biomedical images in MATLAB - A fundamental tool for scientific computing and engineering applications. InTech. 2012 volume 3.

[6] Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965;30:179-185

[7] Bogowicz M, Riesterer O, Ikenberg K, et al. CT radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma Int J Rad Onc Biol Phys. 2017.

[8] Jolliffe IT. Discarding variables in a pricnipal component analysis: artifical data. Journal of the Royla Statistical Society. Series C (Applied Statistics). 1972:21(2):160-173

[9] Cover TM, Thomas JA. Elements of Information Theory. Wiley. 1991

[10] De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics. 2013;29(18):2365-2368

[11] Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. KTK Scientific Publishers. 1986

[12] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2011;73:273-282.

[13] Breiman L. Random forests. Machine Learning 2001:45(1):5-32

Chapter 4

Discussion and outlook

4.1 Main findings

This PhD project aimed to identify imaging phenotypes related to worse prognosis in radiotherapy of head and neck squamous cell carcinoma (HNSCC). Three types of images were analyzed:

- computed tomography (CT) related to density
- computed tomography perfusion (CTP) related to perfusion
- 18F-fludeoxyglucose (18F-FDG) positron emission tomography (18F-FDG PET) related to metabolism.

In the **Chapter 3.1**, I have shown that tumor perfusion depended on the HNSCC subtype and tumor metabolism depended on tumor stage. Moreover, increased blood volume and mean transit time in tumor in comparison to surrounding tissue can be linked to HNSCC subgroups associated with worse treatment outcome (HPV negative and T3/T4 stage tumors). Tumor perfusion, assessed by the CTP, provides additional information in comparison to routinely acquired CT and PET images. However, during this examination a high radiation dose is delivered to a patient. Therefore, in **Chapter 3.2** I have proposed a modified CTP imaging protocol, which enables dose reduction and maintenance of image quality. I have further studied the CTP imaging in context of radiomics analysis (**Chapter 3.4**) and identified 10 groups of CTP radiomic features, which were uncorrelated and robust against CTP calculation factors.

Radiomics is a relatively new, but rapidly growing field of research [17]. Therefore, a large effort is made to provide better understanding of the implemented processes, their impact on models stability as well as the interpretation of radiomics-based prognostic models. In **Chapter 3.3**, I have compered two independent radiomics software implementations in the context of reproducibility of prognostic models. My results have shown, for the first time on a clinically relevant dataset, that radiomics model developed with one implementation cannot be transferred to another radiomics implementation without a detailed comparison. In another study, I have observed an association of CT radiomics with local tumor control and HPV status (**Chapter 3.5**). Tumors more heterogeneous in CT density were linked to worse tumor control and HPV negative status. The CT radiomics model for local tumor control showed a higher prognostic power than a model based on clinical parameters (tumor stage, volume and HPV status). Further, I have investigated an improvement of local tumor control radiomics-based model by inclusion of the 18F-FDG PET imaging (**Chapter 3.6**). Interestingly, the metabolic information assessed by PET imaging did not increase the prognostic power of the model in comparison to CT radiomics. Nonetheless, my radiomics-based local tumor control models showed a higher prognostic power than previously published models for loco-regional control and overall survival [69, 104, 105]. It showed a close link of primary tumor radiomics to a direct endpoint – local control. Consequently, to improve models for the composite endpoints, they should be trained on data from primary tumor radiomics and additional input such as clinical parameters or lymph nodes radiomics.

These main findings are discussed in light of the available literature and future studies directions in the following subsections of the discussion:

- software used for radiomics studies (concerns results in **Chapters 3.3-3.6**)
- challenges of radiomics (concerns results in **Chapters 3.3-3.6**)
- radiomics as a non-invasive imaging biomarker (concerns results in **Chapters 3.5-3.6**)
- heterogeneity of head and neck cancer in terms of imaging phenotypes (concerns results in **Chapters 3.1 and 3.5-3.6**).

4.2 Radiomics software implementation development

The first part of radiomics-related studies was devoted to development of radiomics software implementation. I have developed a stand-alone application written in Python programing language (version 2.7). Development of an own implementation brings an advantage in full understanding of the processes behind the complex calculation and the possibility for further extension of the implementation with new and unique radiomic features. On the other hand, radiomics is a relatively new field of research [17], which lacks standardization. The use of an independent implementation may introduce bias to the trained models. Our group is a part of Image Biomarkers Standardization Initiative (IBIS). It was initiated by Alex Zwanenburg (OncoRay–National Center for Radiation Research in Oncology, Dresden, Germany) and currently involves 19 institutions from 8 countries. It aims to provide clear guidelines for radiomics implementation, from image pre-processing to the calculation of the features [68]. This is an ongoing work and therefore the results are not part of the thesis.

My radiomics application is fully DICOM (Digital Imaging and Communications in Medicine) compatible, which allows for a direct analysis of all medical image modalities without any compression. Currently, the analysis can be performed on CT, PET, CTP and MR, including both clinical and preclinical images. The analysis is performed in a region of interest (ROI) defined in a DICOM structure file, which enables a precise definition of analyzed region. Currently the software encompasses all standard radiomic descriptors:

- shape-based,
- intensity-based,
- texture-based (the Gray-Level Co-Occurrence Matrix, the Neighborhood Gray Tone Difference Matrix, the Gray Level Size Zone Matrix, the Gray Level Run Length Matrix, the Gray Level Distance Zone Matrix and the Neighboring Grey Level Dependence Matrix [68])
- transform-based (wavelet transform).

In total, it provides 1404 radiomic features calculated per image. It enables image discretization into fixed number of bins or fixed bin size. It provides image pre-processing (image and ROI resizing to a cubic voxels), 2D and 3D image analysis. Further details on the implementation, including definition of single features, can be found in **Chapter 3.3** supplement.

4.3 Challenges in radiomics

A large part of radiomics studies is focused on providing proof for radiomics credibility and assuring the reproducibility of trained models. The main challenges in radiomics can be divided into: standardization, robustness against image acquisition, reconstruction and quality, and finally data dimensionality reduction and modeling.

4.3.1 Radiomics standardization

Recently, the Image Biomarker Standardization Initiative (IBIS) provided a compendium of radiomic feature definitions and thus contributed greatly to that field [68]. The compendium standardizes nomenclature and contains a clear description of feature extraction methods for shape, intensity and texture analysis. Most of the radiomics studies are conducted using in-house developed implementations or open source codes [93]. The efforts of IBIS helped to improve the reproducibility of radiomic features between different software implementations [106]. Its current efforts are focused on the standardization of an image pre-processing workflow.

In parallel, I have investigated an influence of different radiomics implementations on the reproducibility of prognostic models **(Chapter 3.3)**. I have compared results from two implementations being a part of IBIS (University Hospital Zurich, Zurich, Switzerland and MAASTRO clinic, Maastricht, the Netherlands). Both implementations were used to train separate local control prognostic models using 3 months post-radiochemotherapy 18-FDG PET scans of patients with head and neck cancer. My final models were reproducible between the implementations, both in the training and validation cohorts. However, in the univariate analysis I have observed that the majority of studied features was not reproducible, despite the fact that only features based on the same definition were selected for this study. I was able to identify steps in the radiomics workflow, which influenced our results (contour mask creation from contour points to voxels, discretization of wavelet transformed images and type of wavelet transform decimated vs undecimated). These steps are rarely described in detail in radiomics papers. Our study highlights the importance of radiomics workflow reporting. Additionally, the high reproducibility of intensity and texture features, despite slightly different contour mask creation methods, indicates lower importance of contour mask creation method in the workflow standardization. Especially in comparison to the interobserver contouring variability, which is one order of magnitude larger [107].

My results also show, for the first time on a clinically relevant model and dataset, that a model developed by one institution should not be directly transferred to another center, which uses a different radiomics implementation, without rigorous comparison.

4.3.2 Radiomics robustness and stability

Higher reproducibility, lower interobserver variability and automatization, which enables the high-throughput analysis, are some of the arguments to favor the quantitative over qualitative image analysis. However, the results of quantitative image analysis can be more sensitive to changes in image quality. Not only the advanced parameters such as texture, but also for example mean SUV depends on voxel size due to partial-volume effect [108]. Table 1 lists some of the factors influencing image quality for the three modalities studied in this project. Some factors are modality-dependent, whereas the others influence multiple modalities. A direct translation of robustness studies between modalities is not trivial. For example, the iodine contrast concentration might influence both contrast enhanced CT and CTP imagining however, the magnitude of difference will be most likely modality-dependent.

Factor influencing image quality	**CT**	**CTP**	**FDG PET**
Reconstruction method	✓	✓	✓
Spatial resolution	✓	✓	✓
Uptake time / scan delay time after contrast administration	✓	✓	✓
Scanner type	✓	✓	✓
Scanner calibration	✓	✓	✓
Blood glucose level			✓
Dose of FDG			✓
Dose of contrast	✓*	✓	
Tube voltage	✓	✓	
Tube current	✓	✓	
Temporal resolution		✓	
Artery input function		✓	
Nosie threshold in the calculation		✓	

*Table 1. Summary of the main factors influencing image quality in computed tomography (CT), computed tomography perfusion (CTP) and fluorodeoxyglucose positron emission tomography (FDG PET); * - only for the contrast enhanced CT.*

Multiple studies indicated voxel size as a parameter having large influence on the stability of radiomic features in different imaging modalities [109, 110]. I have also confirmed this in CT perfusion (CTP) imaging **(Chapter 3.4)**. Recent studies have shown that voxel resampling to a fixed dimensions across a dataset can increase the stability [110, 111]. However, the question as to what is the best voxel dimension requires further investigation. Voxels can be resampled to small dimensions assuring high resolution but introducing a bias from interpolation. On contrary, a lowest resolution available in the dataset can be used, which introduces smoothing and might cause loss of information. Currently it is not clear, which method is preferable for radiomics.

In the analysis of retrospective data, the influence of imaging protocol on image quality and radiomics analysis can be difficult to assess. One of the examples is image reconstruction method. Raw data, which is mostly deleted few days after the examination, are required to compare different reconstruction settings. In the process of image reconstruction, differences in image quality can be caused by for example: the filtered back-projection (FBP) reconstruction vs the iterative reconstruction, a kernel used in the FBP, parameters of iterative reconstruction and 2D vs 3D reconstruction [112-114]. To compare a 2D vs 3D image reconstruction, data have to be additionally acquired in 2D, which is no longer a standard practice and would require an additional approval from an ethics committee. Another example is the scanner type influence on image quality. Both, influence of scanner type and reconstruction method, can be studied using a phantom [110, 115]. However, it may be difficult to reproduce the complex patient anatomy or tumor metabolism in a phantom [115]. Moreover, some of the old scanners may have already been replaced. To overcome this, I have trained my radiomics prognostic models on retrospective data however, the validation was performed in a prospective cohort with a standardized imagining protocol **(Chapters 3.3, 3.5 and 3.6)**. For example, PET data in our training cohort were reconstructed with 2D or 3D iterative algorithm, but in the validation cohort all data were reconstructed with 3D algorithm. Using this approach the balance of the data has to be checked to avoid bias. For my CT signatures, for local tumor control and HPV status, I have shown that the selected features differed between patient groups (local control vs no local control and HPV+ vs HPV-) but not between different scanners **(Chapter 3.5)**. Only two features showed a small dependency on the scanner type, but both signatures were valid in the third scanner type used in the validation cohort. Of course this alternative method does not assure the stability of the models among all possible scanners types and vendors however, it increases their credibility. Training a model on heterogeneous data can also cause an underestimation of the power of the model. Nevertheless, it is of a lower clinical significance than model overfitting to a restricted but not well-documented population.

The quality of CTP is affected by acquisition and computation factors, which are specific for this modality: temporal resolution, arterial input function definition and noise threshold cutoff in the calculation. To my knowledge, my study is the only work in radiomics field trying to quantify the influence of CTP-specific factors on the stability of CTP radiomics **(Chapter 3.4)**. I have analyzed 945 radiomics features in both lung and head and neck cancer patients. I have identified only 255 features stable in respect to non-standardizable CTP calculation factors (arterial input function definition and noise threshold). Within this subset, after the correction for inter-features correlation and correlation with tumor volume, I have identified only 10 independent groups. This greatly reduces the number of CTP radiomics features suitable for prognostic modeling. Additionally, my study showed that stability of the features is tumor site dependent. This was recently confirmed in CT thoracic imaging by Larue et al [116].

The CTP imaging is a dose intense examination. Therefore, I have adapted our clinical protocol to reduce delivered dose and maintain signal to noise ratio. It was done by the application of an iterative reconstruction algorithm as well as tube current and voltage adaptation **(Chapter 3.2)**. Based on a phantom study and measurements using thermoluminescent dosimeters, I have shown that the dose reduction by factor of 2 is feasible without a decrease in image quality. Further investigation of the influence of the new CTP protocol on stability of radiomic features based on patient data is needed.

4.3.3 Outcome modeling and correlation to clinical parameters

Modeling and data dimensionality reduction is a challenge not only in radiomics but also in genomics, proteomics, metabolomics and in general any kind of -omics analysis [117]. To handle a large dimensionality of data and relatively few events machine learning approaches are tested. Pamar et al [118] has shown that different combinations of feature selection and classification methods (training method) resulted in different CT radiomic signatures for overall survival in lung cancer, characterized by different prognostic power with concordance index (CI) ranging from 0.50 (random prediction) to

0.69. The same group investigated the impact of training method on prognostic signature for overall suvival in head and neck cancer [119]. Similarly, they observed a wide range of CI (0.50 – 0.79) dependent on the used method. Additionally, they found that the method providing 'the best' signature, the singature with the highest CI, was dependent on tumor entity. The studies of Pamar et al and Leger et al [104] showed that the performance of trained models was influenced in a larger extent by the feature selection method than the classification method. Recently, Zhang et al [120] studied performance of machine learning classifiers in MRI radiomics in nasopharyngeal cancer. The method providing the best results in MRI radiomics was different than the method found by Pamar et al in CT radiomics in head and neck cancer [119]. Another study showed that the performance of machine learning methods was influenced by the studied endpoint and sample size [121].

To that end, I have investigated few feature selection and classification methods (including traditional statistical methods and machine learning methods) for training local tumor contol models in head and neck cancer usning CT or PET radiomics **(Chapter 3.6)**. I have observed that the same method (principal component analysis combined with multivariable Cox regression) resulted in the best models for both imaging modalities. My results suggest that the imaging modality may not be a critical factor in training method selection. In the future, studies correlating radiomic sigature to clinical endpoint should investigate several methods for model training. Currently, it is not clear which factor has the biggest influence on perofrmnace of those methods: tumor site, sample size, clinical endpoint or image modality.

In conclusion, my work in the field of robustness in radiomics contributed to higher creadibility of radiomics models. I have shown that the models trained on retrospective and heterogenous data can be reproduced on a prospective cohort with a standardized imaging protocol. In light of muliptle workflow steps influencing the output of radiomics software implementation, I recommend, addtionally to thorough reporting of the methodology, an online publication of sample dataset (images together with computed radiomic siganture) to allow for a comparison between implementiations. Finally, I have identified a set of stable CTP radiomic features in respect to perfusion calculation factors. Those features should be tested for their prognostic power in a clinically relevant dataset.

4.4 Radiomics as a non-invasive imaging biomarker

The potential of radiomics as a non-invasive biomarker in head and neck cancer has been recognized [84]. As part of this PhD project, I have trained and validated a CT radiomics based prognostic model for local tumor control in head and neck squamous cell carcinoma (HNSCC) **(Chapter 3.5)**. I have observed that tumors more heterogeneous in CT density are at higher risk of recurrence. These findings are consistent with a previously published CT model for overall survival by Aerts et al. [69]. Additionally, the prognostic power of my model (concordance index CI = 0.78) was higher than the prognostic power of CT radiomics overall survival model (CI = 0.69). This supports the hypothesis presented in **Chapter 2** that primary tumor heterogeneity is not sufficient to describe such a complex endpoint as overall survival [122]. In my training cohort, the prediction of local tumor control only slightly improved when CT radiomics was combined with clinical prognostic factors (tumor stage, volume and HPV status). In the validation cohort CT radiomics showed higher prognostic power than a combined model. Additionally, the CT radiomics model was also prognostic in the subgroup of HPV negative tumors, with a significant split into two risk groups.

In the next study, I have compared the prognostic power of CT and 18F-FDG PET radiomics in the same cohort of HNSCC **(Chapter 3.6)**. I have found, that the round tumors (based on PET signal autosegmentation) with a focused region of high FDG uptake surrounded by a rim of low FDG uptake were associated with better prognosis. In my early study **(Chapter 3.1)**, I have observed that some of the PET measures depended on tumor stage, for example SUV mean. In contrary, my PET radiomics model was not tumor stage dependent. In the comparison to CT radiomics, the radiomics analysis of tumor metabolism using 18F-FDG PET did not improve the prediction of local tumor control. Both models showed similar prognostic power in the validation cohort. Additionally, the combination of CT and PET radiomics did not improve the modeling. Nevertheless, the PET radiomics model showed a better calibration in the validation set in comparison to the CT model. Some of the misclassified patients were affected by CT artifacts in the tumor region, which might have affected the calibration. Further, the analysis of 3 months post-treatment 18F-FDG uptake also did not improve the prediction of local tumor control **(Chapter 3.3)** in comparison to pre-treatment imaging. In that study the radiomics was calculated in the location of primary tumor (the primary tumor contours were transferred to post-treatment imaging using rigid registration). A deformable registration or an independent primary tumor bed definition on post-treatment imaging could be considered in the future to improve the predictions.

In the context of correlation between tumor biology and radiological phenotype, to the best of my knowledge, my study investigating HPV CT radiomics signature was the first one in HNSCC with an independent validation **(Chapter 3.5)**. My results obtained on two independent cohorts of patients, supported the hypothesis from exploratory studies [99, 100] that HPV-positive tumors are more homogenous in CT density.

The CT radiomics signature for HPV status was independent from my CT radiomic signature for local tumor control, despite the fact that they both point to tumor heterogeneity. The local control signature identifies more heterogeneous tumors as a group with worse prognosis. In the same line, the HPV signature indicates that more heterogeneous tumors are HPV negative and this subgroup is known to be linked to worse prognosis. However, these signatures comprised different radiomic features and the predictions obtained from both models were not correlated. For example, the CT radiomics signature for HPV status prediction was valid in the subcohorts of patients with and without local tumor control (**Chapter 3.5** supplement).

In the future, our group plans to further validate the obtained results in external datasets. Especially, the PET radiomics results may be biased by single institution data considering the influence of scanner calibration on the conversion to activity measurements [123]. We also plan to investigate in more detail, the unexpected results suggesting that PET radiomics does not improve the predictions of local tumor control in comparison to CT radiomics. Our results were obtained based on the analysis of autosegmented PET volume. It is well-known that parts of tumors do not exhibit increased metabolism. Possibly, PET radiomics performed on the entire GTV will improve local tumor control predictions. Metal artifacts are a common concern in head and neck cancer imaging. In my CT radiomics projects, I have excluded the CT slices affected by artifacts from the analysis. However, a worse calibration in CT radiomics model vs PET radiomics model correlated with the presence of artifacts in some on the misclassified cases. Possibly, my artifacts exclusion technique biased the CT radiomics results. Future studies investigating the impact of the artifacts on the modeling (percentage of volume affected by artifacts or iterative reconstruction algorithms for artifacts reduction) are needed.

In the context of endpoint selection, I have shown that my CT radiomics local tumor control model is more prognostic than overall survival (OS) model [69] and loco-regional control (LRC) model [104]. Recently, Leger et al have also reported that CT radiomics has a higher prognostic power for prediction of LRC than OS [104]. This observation holds true also for the PET radiomics. I have found a significant association between PET radiomics and local tumor control, whereas Vallières et al have not observed any association with LRC [105]. In the future, we plan to investigate whether the prediction of more complex endpoints can be improved by inclusion of lymph node radiomics to LRC modeling and other non-tumor related factors (overall patient status, alcohol consumption, tobacco abuse and age) to OS modeling.

In conclusion, the results from this PhD project support the importance of radiomics in outcome modeling in HNSCC and indicate the areas for further research in this field. Finally, we plan to continue our efforts to find a link between the radiological tumor phenotype with tumor biology via the correlation of radiomic features with histopathology (cell proliferation, microvessl density or hypoxia).

4.5 Heterogeneity of head and neck cancer and its implication on radiotherapy

HNSCC has long been considered as a uniform disease [124]. The detailed histopathology and genetic studies changed that assumption. The histopathology studies revealed three main variants of head and neck tumors: basaloid, spindle-cell and papillary [124]. HPV positive tumors were found to belong to the basaloid subtype. Additionally, the genetic profile and mutation rate was found to be HNSCC subtype dependent [125]. HPV negative tumors exhibit twice the mutation rate of HPV positive ones. Similarly, laryngeal cancer is characterized by a higher mutation rate than oral cavity, oropharynx or hypopharynx cancers. My imaging-based study correlating perfusion and FDG uptake in the HNSCC, showed its dependency on tumor subtype **(Chapter 3.1)**. Only the subgroup of HPV positive tumors showed positive correlation between mean blood flow and mean SUV. This can explain the inconclusive results published by other groups, where mixed tumor subtypes were analyzed [126-128]. Additionally, I have observed that tumors linked with worse prognosis (advanced stages T3/T4 and HPV negative HNSCC) are characterized with increased blood volume and mean transit time in comparison to the surrounding tissue. This could indicate an imaging-based signature of worse prognosis however, warrants further investigation and validation. Identification and targeting of tumor subtypes is a focus of precision medicine.

In the era of precision medicine, identification of biomarkers, helping to guide treatment decision, is crucial. One of the examples in HNSCC are the results of the DAHANCA 5 study investigating the benefit of radiotherapy combined with nimorazole, a hypoxia modifier [129]. Based on the 15 genes signature for hypoxia they showed that nimorazole increases the LRC rate only in hypoxic tumors. Moreover, no difference in LRC control was observed for HPV positive tumors for two treatment arms irrespective of hypoxia level. Another group proposed a gene signature related to tumor intrinsic radiosensitivity [130]. Biopsy or blood serum based biomarkers are of high interest as they are a direct measure of tumor biology. Medical imaging is another potentially interesting biomarker, although it is mostly only a surrogate of tumor biology. However in contrast to biopsy or blood serum based biomarkers, it does not show limitations in capturing of the spatial and temporal changes in tumor phenotype [131]. In the context of radiation therapy, imaging biomarkers are of high relevance as this treatment option can shape the delivered dose according to 3D distribution of radiosensitivity.

Few clinical trials are currently running, investigating the efficiency of imaging-guided (FDG or FMISO) dose redistribution in HNSCC [132, 133]. This year, a first prognostic radiomic signature in HNSCC was proposed. This signature was able to identify a group of patients receiving cetuximab with LRC as high as in patients receiving cisplatin [134]. However, it requires further validation. I have shown, on the example of HNSCC and HPV status, that detection of biological information is feasible even using CT, a simple morphologic imaging technique. Additionally, I have observed that a radiomic signature based on primary tumor has a higher prognostic power for local tumor control than overall survival. This close link to tumor control should be further tested in the context of intrinsic tumor radiosensitivity. In a future study, parts of the tumor characterized by higher heterogeneity in CT density could be clustered based on a subtumor (regional) radiomics and compared with the location of in field recurrence. If such a correlation could be established, this can lead to a radiomics-based dose painting treatment modification. However, radiomics is a top-down approach and therefore extensive validation of risk group stratification is required before it will be used in clinical trials either as biomarker for treatment intensification or dose redistribution.

In summary, studies performed as part of this PhD project provided an additional evidence of imaging-based inter- and intra-tumor heterogeneity. Imaging characteristics, which may be linked with worse prognosis were identified: increased blood volume and mean transit time in comparison to surrounding tissue, heterogeneity of CT density or aspherical tumors with large subregions of high FDG uptake. However, these characteristics should be further validated in external cohort of patients and treatment intensification options for worse prognosis groups have to be defined.

References

1. Burrell, R.A., et al., *The causes and consequences of genetic heterogeneity in cancer evolution.* Nature, 2013. **501**(7467): p. 338.
2. Vigneswaran, N. and M.D. Williams, *Epidemiological trends in head and neck cancer and aids in diagnosis.* Oral and maxillofacial surgery clinics of North America, 2014. **26**(2): p. 123.
3. Lassen, P., et al., *Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer.* Radiotherapy and Oncology, 2014. **113**(3): p. 310-316.
4. Ang, K.K., et al., *Human papillomavirus and survival of patients with oropharyngeal cancer.* New England Journal of Medicine, 2010. **363**(1): p. 24-35.
5. World Health Organization Regional Office for Europe. *Cancer data and statistics*. 2017 [cited 2017 30 Aug]; Available from: http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/data-and-statistics.
6. World Health Organization. *Cancer key facts*. 2017 [cited 2017 30 Aug]; Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.
7. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2016.* CA: a cancer journal for clinicians, 2016. **66**(1): p. 7-30.
8. Alberts, B., *Molecular biology of the cell*. 2017: Garland science.
9. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* cell, 2011. **144**(5): p. 646-674.
10. Liberti, M.V. and J.W. Locasale, *The Warburg effect: how does it benefit cancer cells?* Trends in biochemical sciences, 2016. **41**(3): p. 211-218.
11. Lamlertthon, W., M.C. Hayward, and D.N. Hayes, *Emerging technologies for improved stratification of cancer patients: a review of opportunities, challenges, and tools.* The Cancer Journal, 2011. **17**(6): p. 451-464.
12. Podgorsak, E., *Radiation oncology physics: a handbook for teachers and students.* Br J Cancer, 2008. **98**: p. 1020.
13. Joiner, M.C. and A. Van der Kogel, *Basic clinical radiobiology*. 2016: CRC press.
14. Tofilon, P.J. and K. Camphausen, *Increasing the Therapeutic Ratio of Radiotherapy*. 2017: Springer.
15. Eschrich, S.A., et al., *A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation.* International Journal of Radiation Oncology* Biology* Physics, 2009. **75**(2): p. 489-496.
16. Fass, L., *Imaging and cancer: a review.* Molecular oncology, 2008. **2**(2): p. 115-152.
17. Lambin, P., et al., *Radiomics: extracting more information from medical images using advanced feature analysis.* European journal of cancer, 2012. **48**(4): p. 441-446.
18. Peyrin, F. and K. Engelke, *CT imaging: Basics and new Trends*, in *Handbook of Particle Detection and Imaging*. 2012, Springer. p. 883-915.
19. Fieselmann, A., et al., *Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details.* Journal of Biomedical Imaging, 2011. **2011**: p. 14.
20. Østergaard, L., et al., *High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis.* Magnetic resonance in medicine, 1996. **36**(5): p. 715-725.
21. Meier, P. and K.L. Zierler, *On the theory of the indicator-dilution method for measurement of blood flow and volume.* Journal of applied physiology, 1954. **6**(12): p. 731-744.
22. Luna, A., et al., *Functional Imaging in Oncology: Biophysical Basis and Technical Approaches*. Vol. 1. 2013: Springer Science & Business Media.
23. Kinehan, P. and J. Fletcher, *PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy.* Semin Ultrasound CT MR, 2010. **31**(6): p. 496-505.

24. Shields, A.F., *Principles of Medical Imaging in the Diagnosis and Staging of the Cancer Patient*, in *Clinical PET-CT in Radiology*. 2011, Springer. p. 147-155.
25. Eubank, W.B., J.H. Lee, and D.A. Mankoff, *Disease restaging and diagnosis of recurrent and metastatic disease following primary therapy with FDG-PET imaging.* PET clinics, 2009. **4**(3): p. 299-312.
26. International Commission on Radiation Units and Measurements, *Report 50: Prescribing, recording and reporting photon beam therapy.* 1993.
27. International Commission on Radiation Units and Measurements, *Report 62: Prescribing, recording and reporting photon beam therapy.* 1999.
28. Buijsen, J., et al., *FDG–PET–CT reduces the interobserver variability in rectal tumor delineation.* Radiotherapy and Oncology, 2012. **102**(3): p. 371-376.
29. Grégoire, V. and A. Chiti, *Molecular imaging in radiotherapy planning for head and neck tumors.* Journal of Nuclear Medicine, 2011. **52**(3): p. 331-334.
30. Nestle, U., S. Kremp, and A.-L. Grosu, *Practical integration of [18 F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): the technical basis, ICRU-target volumes, problems, perspectives.* Radiotherapy and oncology, 2006. **81**(2): p. 209-225.
31. Ling, C.C., et al., *Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality.* International Journal of Radiation Oncology* Biology* Physics, 2000. **47**(3): p. 551-560.
32. Thorwarth, D., X. Geets, and M. Paiusco, *Physical radiotherapy treatment planning based on functional PET/CT data.* Radiotherapy and Oncology, 2010. **96**(3): p. 317-324.
33. Thorwarth, D. and M. Alber, *Implementation of hypoxia imaging into treatment planning and delivery.* Radiotherapy and Oncology, 2010. **97**(2): p. 172-175.
34. Bentzen, S.M. and V. Gregoire. *Molecular imaging–based dose painting: A novel paradigm for radiation therapy prescription*. in *Seminars in radiation oncology*. 2011. Elsevier.
35. Therasse, P., et al., *New guidelines to evaluate the response to treatment in solid tumors.* Journal of the National Cancer Institute, 2000. **92**(3): p. 205-216.
36. Abgral, R., et al., *Does 18F-FDG PET/CT improve the detection of posttreatment recurrence of head and neck squamous cell carcinoma in patients negative for disease on clinical follow-up?* Journal of Nuclear Medicine, 2009. **50**(1): p. 24-29.
37. Gupta, T., et al., *Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis.* European journal of nuclear medicine and molecular imaging, 2011. **38**(11): p. 2083.
38. Galldiks, N., et al., *Role of O-(2-18F-fluoroethyl)-L-tyrosine PET for differentiation of local recurrent brain metastasis from radiation necrosis.* Journal of Nuclear Medicine, 2012. **53**(9): p. 1367-1374.
39. Lordick, F., et al., *PET to assess early metabolic response and to guide treatment of adenocarcinoma of the oesophagogastric junction: the MUNICON phase II trial.* The lancet oncology, 2007. **8**(9): p. 797-805.
40. Nahas, S.C., et al., *Pathologic complete response in rectal cancer: can we detect it? Lessons learned from a proposed randomized trial of watch-and-wait treatment of rectal cancer.* Diseases of the Colon & Rectum, 2016. **59**(4): p. 255-263.
41. O'Sullivan, K., E. Hurley, and J. Hurley, *Understanding complete pathologic response in oesophageal cancer: implications for management and survival.* Gastroenterology research and practice, 2015. **2015**.
42. Mees, G., et al., *Molecular imaging of hypoxia with radiolabelled agents.* European journal of nuclear medicine and molecular imaging, 2009. **36**(10): p. 1674-1686.
43. Dubois, L.J., et al., *Preclinical evaluation and validation of [18F] HX4, a promising hypoxia marker for PET imaging.* Proceedings of the National Academy of Sciences, 2011. **108**(35): p. 14620-14625.
44. Goh, V., et al., *Colorectal tumor vascularity: quantitative assessment with multidetector CT—do tumor perfusion measurements reflect angiogenesis?* Radiology, 2008. **249**(2): p. 510-517.
45. Collins, F.S. and H. Varmus, *A new initiative on precision medicine.* New England Journal of Medicine, 2015. **372**(9): p. 793-795.
46. Zhang, Y. and J. He, *The development of targeted therapy in small cell lung cancer.* Journal of thoracic disease, 2013. **5**(4): p. 538.
47. Moreira, J., et al., *Targeted Therapy in Head and Neck Cancer: An Update on Current Clinical Developments in Epidermal Growth Factor Receptor-Targeted Therapy and Immunotherapies.* Drugs, 2017: p. 1-15.
48. Shin, S.H., A.M. Bode, and Z. Dong, *Precision medicine: the foundation of future cancer therapeutics.* npj Precision Oncology, 2017. **1**(1): p. 12.

49. Collins, D.C., et al., *Towards precision medicine in the clinic: From biomarker discovery to novel therapeutics.* Trends in pharmacological sciences, 2017. **38**(1): p. 25-40.
50. Oldenhuis, C., et al., *Prognostic versus predictive value of biomarkers in oncology.* European Journal of Cancer, 2008. **44**(7): p. 946-953.
51. Ballman, K.V., *Biomarker: predictive or prognostic?* Journal of Clinical Oncology, 2015. **33**(33): p. 3968-3971.
52. Piccart, M., et al., *The predictive value of HER2 in breast cancer.* Oncology, 2001. **61**(Suppl. 2): p. 73-82.
53. O'Connor, J.P., et al., *Imaging biomarker roadmap for cancer studies.* Nature reviews. Clinical oncology, 2017. **14**(3): p. 169.
54. Pulte, D. and H. Brenner, *Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis.* The oncologist, 2010. **15**(9): p. 994-1001.
55. Rusthoven, K.E., D. Raben, and C. Chen, *Improved survival in patients with stage III-IV head and neck cancer treated with radiotherapy as primary local treatment modality.* International Journal of Radiation Oncology* Biology* Physics, 2008. **72**(2): p. 343-350.
56. Rettig, E.M. and G. D'Souza, *Epidemiology of head and neck cancer.* Surgical oncology clinics of North America, 2015. **24**(3): p. 379-396.
57. Grégoire, V., et al., *Squamous cell carcinoma of the head and neck: EHNS–ESMO–ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Annals of Oncology, 2010. **21**(suppl_5): p. v184-v186.
58. Rades, D., et al., *Radiochemotherapy versus surgery plus radio (chemo) therapy for stage T3/T4 larynx and hypopharynx cancer–results of a matched-pair analysis.* European Journal of Cancer, 2011. **47**(18): p. 2729-2734.
59. Schoenfeld, G.O., et al., *Patterns of failure and toxicity after intensity-modulated radiotherapy for head and neck cancer.* International Journal of Radiation Oncology* Biology* Physics, 2008. **71**(2): p. 377-385.
60. Ghosh, G., R. Tallari, and A. Malviya, *Toxicity Profile of IMRT Vs. 3D-CRT in Head and Neck Cancer: A Retrospective Study.* Journal of clinical and diagnostic research: JCDR, 2016. **10**(9): p. XC01.
61. EU Clinical Trials Register. *Clinical trials for head and neck cancer radiotherapy* 2017 [cited 2017 24 Aug].
62. Riesterer, O., L. Milas, and K.K. Ang, *Use of molecular biomarkers for predicting the response to radiotherapy with or without chemotherapy.* Journal of Clinical Oncology, 2007. **25**(26): p. 4075-4083.
63. Kang, H., A. Kiess, and C.H. Chung, *Emerging biomarkers in head and neck cancer in the era of genomics.* Nature reviews Clinical oncology, 2015. **12**(1): p. 11-26.
64. Differding, S., F.-X. Hanin, and V. Grégoire, *PET imaging biomarkers in head and neck cancer.* European journal of nuclear medicine and molecular imaging, 2015. **42**(4): p. 613-622.
65. Pak, K., et al., *Prognostic value of metabolic tumor volume and total lesion glycolysis in head and neck cancer: a systematic review and meta-analysis.* Journal of Nuclear Medicine, 2014. **55**(6): p. 884-890.
66. King, A.D. and H.C. Thoeny, *Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations.* Cancer Imaging, 2016. **16**(1): p. 23.
67. Thorwarth, D., et al., *Combined uptake of [18 F] FDG and [18 F] FMISO correlates with radiation therapy outcome in head-and-neck cancer patients.* Radiotherapy and oncology, 2006. **80**(2): p. 151-156.
68. Zwanenburg, A., et al., *Image biomarker standardisation initiative.* arXiv:1612.07003v4, 2017.
69. Aerts, H.J., et al., *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.* Nature communications, 2014. **5**.
70. Haralick, R., K. Shanmugam, and I. Dinstein, *Textural features for image classification.* IEEE Trans. Syst. Man Cybern, 1973. **3**: p. 610-621.
71. Thibault, G., et al., *Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification.* Pattern Recognition Inf Process., 2009: p. 140-145.
72. Solomon, C. and T. Breckon, *Image Segmentation, in Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. 2010, Chichester, UK. : John Wiley & Sons, Ltd.
73. Broughton, S. and K. Bryan, *Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing*. 2008, Hoboken, New Jersey: John Wiley & Sons, Inc.
74. Coroller , T., et al., *CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma.* Radiother Oncol., 2015. **114**(3): p. 345-350.

75. Ha, S., et al., *Metabolic Radiomics for Pretreatment 18F-FDG PET/CT to Characterize Locally Advanced Breast Cancer: Histopathologic Characteristics, Response to Neoadjuvant Chemotherapy, and Prognosis.* Sci Rep. 2017, 2017. **7**(1): p. 1556.
76. Huang, Y., et al., *Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer.* Radiology., 2016. **281**(3): p. 947-957.
77. Kickingereder, P., et al., *Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response.* Clin Cancer Res. , 2016. **22**(23): p. 5765-5771.
78. Lee, G., et al., *Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art.* Eur J Radiol., 2017. **86**: p. 297-307.
79. Leijenaar, R., et al., *External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma.* Acta Oncol., 2015. **54**(9): p. 1423-1429.
80. Parmar, C., et al., *Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer.* Sci Rep., 2015. **5**: p. 11044.
81. Scalco, E. and G. Rizzo, *Texture analysis of medical images for radiotherapy applications.* Br J Radiol., 2017. **90**(1070): p. 20160642.
82. Scrivener, M., E. de Jong, and J. van Timmeren, *Radiomics applied to lung cancer: a review.* TCR, 2016. **5**(4).
83. Vallières, M., et al., *A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities.* Physics in medicine and biology, 2015. **60**(14): p. 5471.
84. Wong, A., et al., *Radiomics in head and neck cancer: from exploration to application.* Translational Cancer Research, 2016. **5**(4).
85. Grossmann , P., et al., *Quantitative Imaging Biomarkers for Risk Stratification of Patients with Recurrent Glioblastoma Treated with Bevacizumab.* Neuro Oncol., 2017: p. in press.
86. Hsu, C., et al., *Tumor compactness improves the preoperative volumetry-based prediction of the pathological complete response of rectal cancer after preoperative concurrent chemoradiotherapy.* Oncotarget., 2017. **8**(5): p. 7921-7934.
87. Rios-Velazquez, E., et al., *Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer.* Cancer Res. 2017, 2017. **77**(14): p. 3922-3930.
88. Yip, S., et al., *Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer.* J Nucl Med. , 2017. **58**(4): p. 569-576.
89. Li, Y., et al., *MRI features can predict EGFR expression in lower grade gliomas: A voxel-based radiomic analysis.* Eur Radiol., 2017: p. in press.
90. Yu, J., et al., *Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma.* Eur Radiol. 2017, 2017. **27**(8): p. 3509-3522.
91. Panth, K.M., et al., *Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells.* Radiotherapy and Oncology, 2015. **116**(3): p. 462-466.
92. Hatt, M., et al., *Characterization of PET/CT images using texture analysis: the past, the present… any future?* European journal of nuclear medicine and molecular imaging, 2017. **44**(1): p. 151-165.
93. Larue, R.T., et al., *Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures.* The British journal of radiology, 2017. **90**(1070): p. 20160665.
94. Sollini, M., et al., *PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology.* Scientific Reports, 2017. **7**(1): p. 358.
95. Leijenaar, R.T., et al., *The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis.* Scientific reports, 2015. **5**.
96. Fave, X., et al., *Computational resources for radiomics.* Translational Cancer Research, 2016. **5**(4): p. 340-348.
97. Sørensen, B.S., et al., *Radiosensitivity and effect of hypoxia in HPV positive head and neck cancer cells.* Radiotherapy and oncology, 2013. **108**(3): p. 500-505.
98. Wistar, C., *A System of Anatomy: For the Use of Students of Medicine. 9th ed.* 1846, Philadelphia: Thomas, Cowperthwait & Co.; 1846.
99. Buch, K., et al., *Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on CT.* American Journal of Neuroradiology, 2015. **36**(7): p. 1343-1348.
100. Fujita, A., et al., *Difference between HPV-positive and HPV-negative non-oropharyngeal head and neck cancer: texture analysis features on CT.* Journal of computer assisted tomography, 2016. **40**(1): p. 43-47.

101. Lambin, P., et al., *Radiomics: the bridge between medical imaging and personalized medicine.* Nat Rev Clin Oncol 2017. **In Press.**
102. Harrison, L.B., R.B. Sessions, and W.K. Hong, *Head and neck cancer: a multidisciplinary approach*. 2009: Lippincott Williams & Wilkins.
103. Paidpally, V., et al., *FDG-PET/CT imaging biomarkers in head and neck squamous cell carcinoma.* Imaging in medicine, 2012. **4**(6): p. 633.
104. Leger, S., et al., *A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling.* Scientific Reports, 2017. **7**: p. 13206.
105. Vallières, M., et al., *Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer.* Sci Rep. , 2017. **7**(1): p. 10117.
106. Zwanenburg, A. and et al., *EP-1677: Multicentre initiative for standardisation of image biomarkers.* Radiotherapy and Oncology, 2017.
107. Riegel, A.C., et al., *Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion.* International Journal of Radiation Oncology* Biology* Physics, 2006. **65**(3): p. 726-732.
108. Hatt, M., et al., *Impact of partial-volume effect correction on the predictive and prognostic value of baseline 18F-FDG PET images in esophageal cancer.* Journal of Nuclear Medicine, 2012. **53**(1): p. 12-20.
109. Galavis, P.E., et al., *Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters.* Acta Oncologica, 2010. **49**(7): p. 1012-1016.
110. Larue, R.T., et al., *Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study.* Acta Oncologica, 2017: p. 1-10.
111. Shafiq-ul-Hassan, M., et al., *Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels.* Medical physics, 2017. **44**(3): p. 1050-1062.
112. van Velden, F.H., et al., *Repeatability of radiomic features in non-small-cell lung cancer [18F] FDG-PET/CT studies: impact of reconstruction and delineation.* Molecular imaging and biology, 2016. **18**(5): p. 788-795.
113. Kim, H., et al., *Impact of reconstruction algorithms on CT radiomic features of pulmonary tumors: analysis of intra-and inter-reader variability and inter-reconstruction algorithm variability.* PloS one, 2016. **11**(10): p. e0164924.
114. Zhao, B., et al., *Reproducibility of radiomics for deciphering tumor phenotype with imaging.* Scientific reports, 2016. **6**: p. 23428.
115. Shiri, I., et al., *The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies.* European Radiology, 2017: p. 1-12.
116. Larue, R.T., et al., *4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers.* Radiotherapy and Oncology, 2017.
117. Alyass, A., M. Turcotte, and D. Meyre, *From big data analysis to personalized medicine for all: challenges and opportunities.* BMC medical genomics, 2015. **8**(1): p. 33.
118. Parmar, C., et al., *Machine learning methods for quantitative radiomic biomarkers.* Scientific reports, 2015. **5**.
119. Parmar, C., et al., *Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer.* Frontiers in oncology, 2015. **5**.
120. Zhang, B., et al., *Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Advanced Nasopharyngeal Carcinoma.* Cancer Letters, 2017.
121. Zhang, Y., et al., *Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer.* Scientific Reports, 2017. **7**.
122. Leoncini, E., et al., *Clinical features and prognostic factors in patients with head and neck cancer: Results from a multicentric study.* Cancer epidemiology, 2015. **39**(3): p. 367-374.
123. Lockhart, C.M., et al., *Quantifying and reducing the effect of calibration error on variability of PET/CT standardized uptake value measurements.* Journal of Nuclear Medicine, 2011. **52**(2): p. 218-224.
124. Pai, S.I. and W.H. Westra, *Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment.* Annual Review of Pathological Mechanical Disease, 2009. **4**: p. 49-70.
125. Stransky, N., et al., *The mutational landscape of head and neck squamous cell carcinoma.* Science, 2011. **333**(6046): p. 1157-1160.
126. Bisdas, S., K. Spicer, and Z. Rumboldt, *Whole-tumor perfusion CT parameters and glucose metabolism measurements in head and neck squamous cell carcinomas: a pilot study using combined positron-emission tomography/CT imaging.* American Journal of Neuroradiology, 2008. **29**(7): p. 1376-1381.

127. Hirasawa, S., et al., *Inverse correlation between tumor perfusion and glucose uptake in human head and neck tumors.* Academic radiology, 2007. **14**(3): p. 312-318.
128. Veit-Haibach, P., et al., *Combined PET/CT-perfusion in patients with head and neck cancers.* European radiology, 2013. **23**(1): p. 163-173.
129. Toustrup, K., et al., *Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in squamous cell carcinomas of the head and neck.* Radiotherapy and Oncology, 2012. **102**(1): p. 122-129.
130. Scott, J.G., et al., *A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study.* The Lancet Oncology, 2017. **18**(2): p. 202-211.
131. Baumann, M., et al., *Radiation oncology in the era of precision medicine.* Nature Reviews Cancer, 2016. **16**(4): p. 234-250.
132. Clinical Trials. *18F-FDG-PET Guided Dose-Painting With Intensity Modulated Radiotherapy in Oropharyngeal Tumours (FiGaRO)*. 2016 [cited 2017 19 Sept, 2017]; Available from: https://clinicaltrials.gov/ct2/show/NCT02953197.
133. Welz, S., et al., *Prognostic value of dynamic hypoxia PET in head and neck cancer: Results from a planned interim analysis of a randomized phase II hypoxia-image guided dose escalation trial.* Radiotherapy and Oncology, 2017.
134. Ou, D., et al., *Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to Human Papillomavirus status.* Oral Oncology, 2017. **71**: p. 150-155.

Curriculum vitae

BOGOWICZ Marta

Birthdate: 22 May 1990
Nationality: Polish

Contact information

Address:
Tramstrasse 79
8050 Zurich, Switzerland

Phone: +41 79 629 05 73
E-mail: marta.bogowicz@usz.ch

Education

Qualification	**Institution**	**Year**
PhD student (cancer biology PhD program)	Faculty of Science, University of Zurich, Switzerland	2014 - now
Medical physics trainee	University Hospital Zurich, Switzerland	2014 - now
MSc cum laude (applications of physics in biology and medicine, specialty: medical physics) Thesis: *Radiobiology model based optimization of dose distribution in radiotherapy.*	Faculty of Physics, University of Warsaw, Poland	2012 - 2014
BSc cum laude (applications of physics in biology and medicine, specialty: medical physics) Thesis: *Beam angles optimization in radiotherapy for patients with lung cancer*.	Faculty of Physics, University of Warsaw, Poland	2009 - 2012
Matura certification	XXVII High School Tadeusz Czacki, Warsaw, Poland	2006 - 2009

Publications

Bogowicz M, Leijenaar RTH, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, Unkelbach J, Guckenberger M, Konukoglu E, Lambin P. *Post-radiochemotherapy PET radiomics in head and neck cancer - the influence of radiomics implementation on the reproducibility of local control tumor models.* submitted to Radiother Oncol.

Stüssi A*, Bogowicz M*, Weichselbaumer V, Veit-Haibach P, Riesterer O, Guckenberger M, Tanadini-Lang S. *Optimizing a Perfusion CT Protocol for Head and Neck Cancer*, Current Directions in Biomedical Engineering 2017, in press

Bogowicz M, Riesterer O, Stark LS, Studer G, Unkelbach J, Guckenberger M, Tanadini-Lang S. *Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma.* Acta Oncol. 2017, in press

Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, Guckenberger M, Tanadini-Lang S. *CT radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma*. Int J Rad Onc Biol Phys. 2017, in press

Bogowicz M, Riesterer O, Bundschuh RA, Veit-Haibach P, Hüllner M, Studer G, Stieb S, Glatz S, Pruschy M, Guckenberger M, Tanadini-Lang S. *Stability of radiomic features in CT perfusion maps.* Phys Med Biol. 2016;61(24)

Nesteruk M, Lang S, Veit-Haibach P, Studer G, Stieb S, Glatz S, Hemmatazad H, Ikenberg K, Huber G, Pruschy M, Guckenberger M, Klöck S, Riesterer O. *Tumor stage, tumor site and HPV dependent correlation of perfusion CT parameters and [18F]-FDG uptake in head and neck squamous cell carcinoma*. Radiother Oncol. 2015;117(1)

*authors contributed equally

Grants and awards

2017 Acta Oncologica travel grant

2016 *The best presentation award* at the 2nd joint meeting of Scientific Association of Swiss Radiation Oncology and Swiss Society of Radiobiology and Medical Physics for presentation: *Radiomics of CT perfusion maps*, Sursee, Switzerland

Selected conference contributions

2017 BiGART 2017, Aarhus, Denmark, June 14-16, 2017. Oral presentation (Bogowicz M *et al.*): *Comparison of PET and CT radiomics for prediction of local tumor recurrence in head and neck squamous cell carcinoma.*

ESTRO 33, Vienna, Austria, May 5-9, 2017. Oral presentation (Bogowicz M *et al.*): *Influence of PET radiomics implementation on reproducibility of tumor control prognostic models*.

2016 ASTRO annual meeting, Boston, the USA, September 25-28, 2016. Oral presentation (Riesterer O, Nesteruk M, Studer G, Guckenberger M, Lang S): *Predictive value of radiomics analysis for local tumor control after radiochemotherapy in patients with head and neck cancer*.

AAPM 58th annual meeting, July 31 – August 4, 2016, Washington DC, the USA. Poster (*Nesteruk M et al.*): *Radiomics in CT perfusion maps of head and neck cancer*.

ESTRO 35, Turin, Italy, April 29 – May 3, 2016. Poster (Nesteruk M et al.): *Radiomics in CT perfusion maps – robustness study*.

2015 3rd ESTRO Forum, Barcelona, Spain, April 24 – 28, 2015. Oral presentation (Nesteruk M *et al.*): *Comparison of perfusion CT parameters and [18F]-FDG uptake in head and neck cancer patients*.

Teaching experience

2016 Teaching assistant at the University of Zurich (programming for biology students) and co-supervisor of semester and master thesis in the field of radiomics.

Acknowledgments

I would especially like to thank my direct supervisor Stephanie Tanadini-Lang for suggesting me this great research topic and for her continuous support. Many thanks to my main supervisor Martin Pruschy for his guidance in the world of biology.

I would like to thank my PhD committee (Martin Pruschy, Matthias Guckenberger, Oliver Riesterer, Uwe Schneider, Stephanie Tanadini-Lang and Jan Unkelbach) for more than 3 years of support and guidance. I would also like to thank Achal Sonbadhar and Diem Voung for language corrections in the thesis.

I would like to thank Xaver Wümrs for his input into shape descriptors in the radiomics implementation. Many thanks to Ralph Leijenaar for endless discussions on wavelet transform.

A sincere thanks to the radiomics group and the whole medical physics group of the Department of Radiation Oncology at the University Hospital Zurich – it was a pleasure to come to work everyday.

Najważniejsze, serdeczne podziekowania dla mojej rodziny (rodzicow Katarzyny i Jozefa oraz siostry Moniki) za ich nieusatjace wsparcie.